



Project  
**MUSE**<sup>®</sup>

*Today's Research. Tomorrow's Inspiration.*

# The Position of the Languages of Eastern Indonesia: A Reply to Donohue and Grimes

Robert Blust

UNIVERSITY OF HAWAI‘I AT MĀNOA

Donohue and Grimes (2008) question the validity of the claims (1) that most of the Austronesian languages of eastern Indonesia fall into a Central Malayo-Polynesian (CMP) subgroup that appears to continue a prehistoric dialect chain, and (2) that Central Malayo-Polynesian and Eastern Malayo-Polynesian form a larger Central-Eastern Malayo-Polynesian (CEMP) subgroup. Some of their objections are valid and welcome, but most of the counterarguments they present are based on an approach to the data that diverges sharply from the standard use of the Comparative Method of linguistics, and in certain crucial cases from generally accepted scientific method. A reexamination of individual comparisons shows that many of the exclusively shared innovations that they dismiss stand up to close scrutiny. This is particularly true of CEMP, and while it is not true of CMP, the evidence for geographically extensive diffusion among the CMP languages strongly suggests that all or most of them once formed a dialect chain.

**1. INTRODUCTION.** Until Dempwolff (1920), no subgrouping claims based on linguistic evidence had ever been advanced for the Austronesian (AN) language family. Well-defined groups such as Polynesian were recognized more on an intuitive basis than through an appeal to exclusively shared innovations. The first real breakthrough in empirically motivated classification came in 1920, when Dempwolff adumbrated the large and important Oceanic group by recognizing the distribution of AN languages that show merger of original \*p and \*b. What is now known as the Oceanic hypothesis was fleshed out in several other publications, culminating in Dempwolff (1934–38), although gaps in documentation left the western boundary of this group unclear.

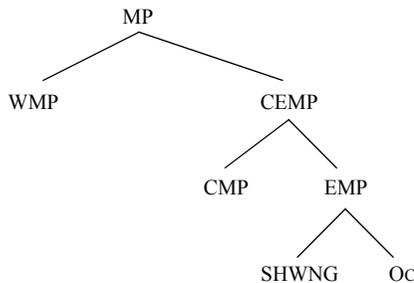
The reality of Oceanic was further strengthened by studies spanning the next thirty years or so, but no other significant work on AN subgrouping appeared until Dyen (1965). This lexicostatistical classification of 245 languages was the largest for any language family at that time, and the first comprehensive phylogenetic tree of the entire AN language family. Unfortunately, however, the methodology used led to massive distortion of actual relationships, pointing to an AN homeland in the region of New Guinea and the Bismarck Archipelago, and thoroughly obscuring the reality of the Oceanic branch.

Little additional work on higher-level subgrouping was done until Blust (1974), where it was suggested on the basis of preliminary evidence that the South Halmahera-West New Guinea language group, which had long been recognized impressionistically in Dutch publications, (1) is a valid genetic group, (2) is the closest relative of Oceanic, and (3) may be part of a larger group that includes most or all AN languages of eastern Indonesia. The reality of South Halmahera-West New Guinea was further strengthened and refined in Blust (1978a), although Dyen (1978) argued that the languages of eastern Indonesia as a whole are no more closely related to Oceanic than they are to the languages of western Indonesia and the Philippines. Blust (1983–84) and (1993) returned to the question of the position of the languages of eastern Indonesia, and reached two conclusions: (1) there is clear evidence for a discrete Central-Eastern Malayo-Polynesian group that includes all languages of eastern Indonesia together with Oceanic, and (2) there is weaker evidence for a Central Malayo-Polynesian group covering much of eastern Indonesia south of Halmahera that probably derives from an earlier dialect chain. To the extent that the resulting tree has been accepted by most scholars concerned with subgrouping issues in the AN language family as well as by scholars who do not work directly with these languages, such as Ruhlen (1987), Campbell (2004:193), and Gordon (2005), it can be called the “standard theory” of Malayo-Polynesian subgrouping. Donohue and Grimes (2008:115) refer to it somewhat disparagingly as “the tree popularly assumed for the AN languages external to Formosa.” The relevant part of this tree is shown in figure 1, where WMP = Western Malayo-Polynesian, CEMP = Central-Eastern Malayo-Polynesian, CMP = Central Malayo-Polynesian, EMP = Eastern Malayo-Polynesian, SHWNG = South Halmahera-West New Guinea, and OC = Oceanic.

Donohue and Grimes (2008:115; hereafter D&G), who use the terms “CEMP area” and “CMP area” to allow noncommittal reference to these proposed groups, object to this tree for reasons that are worth quoting *in extenso*:<sup>1</sup>

We argue that because the few apparent innovations occur only sporadically within the CEMP area, the presence of these innovations in languages that are geographically contiguous to the CEMP-area languages suggests that any

**FIGURE 1. THE “STANDARD THEORY” OF MALAYO-POLYNESIAN BRANCHING**



1. Among other descriptive conventions that D&G employ is the use of modern political units to refer to the area in question (“Eastern Indonesia and East Timor”). I adhere instead to traditional geographical labels.

PCEMP node should, logically, include some languages that have been assumed to be classified as “WMP.” We conclude that, when considering the status of these innovations with respect to the AN languages to their west, the putatively defining isoglosses do not add up to a subgrouping argument.

With regard to the CMP hypothesis they continue:

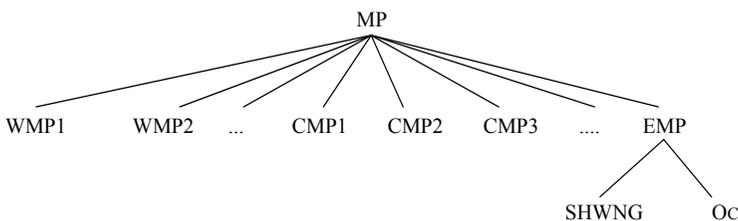
The fact that Blust claims to have reconstructed a linkage, rather than an undifferentiated protolanguage, means that we should expect some degree of incompleteness in the attestation of these forms, but that the incompleteness should be geographically defined, rather than random. The fact that we find many of the proposed innovations outside the CMP area, in the WMP area, argues for a greater fragmentation than can be assumed with a linkage that is still historically and geographically limited.

In place of the “standard tree” they propose the tree in figure 2 as “a possible reflection of the data.”

D&G (2008:117) state that “the history of the CMP and CEMP proposals has been one of quiet acceptance in print, with the exception of Ross (1995); Grimes, Therik, Grimes, and Jakob (1997); and Adelaar (2005a).” Given their critique, this makes four published attacks on these proposals in the span of just fifteen years, hardly what most readers would call “quiet acceptance.” Add to that the conference papers of Grimes (1991) and Nothofer (1992), and one might easily conclude that the CMP and CEMP hypotheses have been under attack from the very start—in fact, even before the publication date of the paper in which they were first fully expressed. As will be noted again below, this is in striking contrast to the reception of Ross’s (1988) proposals for higher-order subgroups in Oceanic, several of which involve collections of languages that are defined by overlapping sets of innovations, and are therefore no more discrete than CMP.

In the remainder of this paper I will re-examine the evidence for the relevant parts of the tree in figure 1, namely for CEMP and CMP. My approach will be to first cite supporting evidence for this tree that I feel has been unjustifiably dismissed, and then to consider some of the alternatives that D&G believe support the tree in figure 2. The approach adopted by D&G in arguing against the CEMP and CMP hypotheses departs significantly from fundamental principles of historical linguistics, and in some cases violates fundamental principles of scientific argumentation in general. For this reason, it is necessary to briefly address some issues of method before examining disagreements in the interpretation of particular linguistic comparisons.

**FIGURE 2. MALAYO-POLYNESIAN BRANCHING  
ACCORDING TO DONOHUE AND GRIMES (2008:116)**



**2. METHODOLOGICAL PRELIMINARIES.** Linguistic subgroups are based on exclusively shared innovations that are not likely to be due to convergence. Innovations fall into two formally distinct types: replacement innovations (schematically \*A > B), and nonreplacement innovations (schematically \*A > A ~ B). With regard to lexical innovations, we must further recognize indeterminate cases in which a form-meaning match is exclusively shared by a subset of languages, but no higher-level reconstruction for this meaning is known.<sup>2</sup> If, for purposes of discussion, we restrict ourselves to the lexicon, replacement innovations can be illustrated by PAN \*aNak > PEMP \*natu ‘child’; the inherited PAN form \*aNak is widely reflected in Taiwan, the Philippines, western Indonesia, and eastern Indonesia up to southern Halmahera, but has disappeared from all EMP languages, being replaced by a great variety of terms, only one of which is shared between SHWNG and OC languages. A nonreplacement innovation in lexicon can be illustrated by PMP \*buRaw ‘chase away, expel’ > Proto-Philippines \*búRaw ~ \*ábuR ‘chase away, expel’ (Blust 1995). Here, known reflexes of \*ábuR are confined to Philippine languages and so can be taken as evidence for a Philippine subgroup, but this form did not replace PMP \*buRaw in Proto-Philippines. An indeterminate innovation can be illustrated by POC \*qebal ‘pandanus mat, possibly sleeping mat’ (Ross, Pawley, and Osmond 1998:80). Reflexes of this form are attested in Oceanic languages from Numbami in the Huon Peninsula region of New Guinea to the Polynesian languages, and so can be taken as one of many pieces of evidence for an Oceanic subgroup, but since a higher-level reconstruction in this meaning is unknown, the status of \*qebal as an innovation in POC remains uncertain.

It is generally agreed that replacement innovations provide the strongest type of evidence for linguistic subgroups, and wherever these can be found, they must be given considerable weight in determining the structure of family trees. However, as already noted, all exclusively shared innovations carry some weight as subgrouping evidence. Just as with common types of sound change, any given nonreplacement or indeterminate innovation in lexicon may not be weighty by itself, but in combination with others it can form part of a substantial body of evidence for subgrouping hypotheses.

As will be seen below, D&G implicitly reject nonreplacement innovations as evidence for subgrouping. Although they are not explicit about the point, the form of a number of their arguments implies that a linguistic element at stage 1 must be completely replaced at stage 2 to constitute an innovation. This is an interpretation of structural change in language that is at variance with the practice of most historical linguists. If a language had a feature \*A before change took place, but A ~ B after that change, the two historical stages are clearly different, and this difference is a by-product of change. That, essentially, is the definition of “innovation” in historical linguistics. D&G do not explain the logical basis for their objection to the use of nonreplacement innovations, but it clearly does not draw on the logic of the Comparative Method.

Second, D&G sometimes abandon the Comparative Method in their search for counterevidence to the CEMP and CMP hypotheses. In several cases that are discussed

2. Protolanguage abbreviations are PAN, Proto-Austronesian; PCEMP, Proto-Central-Eastern Malayo-Polynesian; PCMP, Proto-Central Malayo-Polynesian; PPH, Proto-Philippines; PMP, Proto-Malayo-Polynesian; POC, Proto-Oceanic; PSHWNG, Proto-South Halmahera-West New Guinea; PWMP, Proto-Western Malayo-Polynesian.

below, as the proposed changes \**kepkep* > Buru *gepe* ‘hold tightly in hand’, and \**likud* > Palu’e *kuri* ‘back’, cognate decisions are based on general phonetic similarity rather than recurrent sound correspondences, leading to false etymologies and flawed conclusions.

Third, when confronted with clear counterevidence to their proposals, as with the reflexes of \**kandoRa* ‘cuscus’, and \**mans(aə)r* ‘bandicoot’, D&G resort to *deus ex machina* explanations, as in proposing recurrent parallel borrowing from an unattested Papuan lingua franca that was once widely used for trading operations in pre-Austronesian Indonesia. As will be detailed below, evasive “explanations” of this kind represent such a flagrant violation of sound scientific method that they justify a reminder of what is meant by an “ad hoc” hypothesis.

Fourth, D&G repeatedly reject comparisons as subgrouping evidence because they are not found in all languages within the proposed group. But, as anyone with experience in subgrouping languages knows all too well, the adoption of such a restriction would stymie all efforts at inferring genetic groupings. The forms \**kiajo* ‘outrigger boom’, \**pakiwak* ‘shark’, and \**soje* ‘famine’ are widely regarded as innovations of POC, but they have a scattered distribution in OC languages. Should they therefore be dismissed as subgrouping evidence, since they are not found in all languages that have been assigned to the OC group? One might argue that Oceanic is already defined by distinctive phonological innovations, and lexical innovations simply provide further confirmation, whereas this is not true of CEMP or CMP, but this is more a matter of degree than of fundamental difference.

Finally, D&G (2008:114, fn. 2) cite two sources for comparative data on AN languages: the Austronesian Basic Vocabulary Database (Greenhill, Blust, and Gray 2005–08), and the Comparative Austronesian Dictionary (Tryon 1995). Surprisingly, they do not refer at all to the Austronesian Comparative Dictionary (Blust 1995), a resource that has been available upon request since at least 1995 and which, despite its incomplete state, provides by far the most complete comparative coverage of the AN lexicon available to date.

**3. CENTRAL–EASTERN MALAYO-POLYNESIAN REVISITED.** D&G (2008:116) open their attack with a gambit that relieves them of the need to consider some of the most crucial evidence presented in Blust (1993): “In this article we address the innovations that appear in Blust (1993), described by him as being ‘phonological’ (both regular and irregular) or ‘semantic’. The lexical innovations are the subject of a separate study (Grimes and Donohue n.d.), and the morphosyntactic innovations can all be shown to be the result of contact-induced change.” The “separate study” to which they refer is a publication that evidently exists only in the planning stages, and the reference to “contact-induced change” draws on several publications by Donohue that advance the novel theory that a Papuan-based trading network spanned the Malay archipelago prior to the arrival of AN speakers during the second millennium BC. In any case, the lexical evidence for CEMP is critically important, and cannot be ignored or postponed for later consideration in any serious critique of this hypothesized subgroup, which includes all AN languages in eastern Indonesia and the Pacific except Chamorro and Palauan. I begin, then, with a review of the lexical evidence for CEMP before considering the criticisms that D&G advance against other types of evidence.

**3.1 MARSUPIAL TERMS.** A key piece of evidence for CEMP comes from cognate sets for the cuscus and bandicoot, two marsupial mammals that were unknown to AN speakers before they crossed the Wallace Line (Blust 1982). Both of these sets are distributed across languages that have been assigned to the CMP, SHWNG and OC languages. Because of their importance in determining the position of the languages of eastern Indonesia, these terms are repeated in (1) and (2) along with cognates from several languages that were unknown at the time of the original formulation (Seimat, Bipi, Nauna, Penchal, Lenkau, and Nauna reflect Proto-Admiralty \*kodaya, with metathesis of V<sub>1</sub> and V<sub>2</sub>).

- (1) PCEMP \*kandoRa ‘cuscus, phalanger’
- |        |                        |                      |
|--------|------------------------|----------------------|
| CMP:   | Watubela               | kadola               |
|        | Kesui (Keldor village) | udora                |
| SHWNG: | Buli                   | do ‘small marsupial’ |
| OC:    | Manam                  | ʔodora               |
|        | Vitu                   | hadora               |
|        | Sori                   | ohay                 |
|        | Seimat                 | koxa                 |
|        | Bipi                   | koxa                 |
|        | Nauna                  | kocay                |
|        | Penchal                | kotay                |
|        | Lou                    | ŋora                 |
|        | Lenkau                 | ŋohay                |
|        | Mussau                 | aroa                 |
|        | Lungga                 | yandora              |
|        | Nggela                 | kandora              |
- (2) PCEMP \*mansar/mansər ‘bandicoot’
- |      |              |                                 |
|------|--------------|---------------------------------|
| CMP: | Leti-Moa     | mada, made                      |
|      | Damar        | madar                           |
|      | Yamdena      | mande ‘cuscus’                  |
|      | Ngaibor      | medar                           |
|      | Ujir         | meday                           |
|      | Kei          | medar ‘cuscus’                  |
|      | Elat (Kei)   | mender                          |
|      | Amblau       | mate                            |
|      | Asisilu      | marel                           |
|      | Alune        | marele                          |
|      | Kamarian     | maker                           |
| OC:  | Motu         | mada                            |
|      | Takia        | madal                           |
|      | Wogeo        | m <sup>w</sup> a <sup>j</sup> a |
|      | Mbula        | moozo                           |
|      | Duke of York | man                             |
|      | Likum        | mucay                           |
|      | Lou          | m <sup>w</sup> as               |
|      | Lenkau       | m <sup>w</sup> as               |
|      | Penchal      | m <sup>w</sup> ac               |
|      | Nauna        | m <sup>w</sup> ac               |
|      | Fijian       | ŋaco <sup>3</sup>               |

These etymologies are important for several reasons. First, although the bandicoot has a more restricted distribution, the cuscus or phalanger is found throughout Sulawesi, and from Timor to the Solomon Islands (Darlington 1980:374). Second, reflexes of \*kandoRa and \*mans(aə)ɾ are found from at least the Leti-Moa archipelago just east of Timor, to the central Solomons. Third, as already noted, if AN languages originated west of the Wallace Line, these forms must be innovations that were acquired when AN speakers entered eastern Indonesia. Fourth, \*kandoRa contains \*o, a vowel that was not present in PAN or PMP. Fifth, although there are irregularities in sound correspondences that will be noted below, these do not prevent an unambiguous reconstruction of protoforms, nor do they support a tenable hypothesis of borrowing. D&G (2008:fn. 6) are not impressed with the subgrouping inference that Blust (1982) draws from these comparisons, holding that the forms in question “could be (and likely are) borrowed from an as yet unidentified non-AN source. Once established in any trade vernacular, they would be spread through all and any subgroups in contact.”

There are at least three serious problems with the alternative that D&G propose:

- (1) No Papuan language in the relevant area is known to have terms related to \*kandoRa or \*mans(aə)ɾ.<sup>4</sup>
- (2) There is no known evidence for a Papuan trade language that was widely used in the Indonesian archipelago before the arrival of AN speakers.
- (3) The subgrouping that D&G propose in figure 2 clearly implies that the AN penetration of eastern Indonesia was the result of at least four historically separate events.

While (1) might be attributed to poor documentation, lexical materials are now available for a number of the Papuan languages of eastern Indonesia and west New Guinea, and since both the cuscus and the bandicoot are common animals in this region, it would be surprising if terms for these animals were not included in most dictionaries or longer wordlists. A preliminary inspection of several dozen sources has turned up no connections. In short, the claim in (2), that there was a Papuan trade language from which \*kandoRa and \*mans(aə)ɾ were borrowed, is without empirical foundation, and appears to be motivated only by its utility in dismissing these lexical innovations as relevant to subgrouping. Finally, if (3) were true, it is difficult to see how terms borrowed from different source languages into different recipient languages would exhibit sound correspondences that would allow reconstructions such as \*kandoRa and \*mans(aə)ɾ. Even if we adopt the unsupported hypothesis that a Papuan trade language was widely used in eastern Indonesia at the time AN speakers arrived, and that \*kandoRa and \*mans(aə)ɾ were borrowed from this hypothetical source, there is no reason to expect

3. Capell (1968:66) glosses this as ‘a variety of large rat, the food of the *mataqali bati*’, where the *mataqali* is a primary social division in traditional Fijian society. Paul Geraghty (pers. comm. 11/17/08) informs me that in the Waidina dialect of southeast Vitilevu *ŋ<sup>v</sup>aco* is a ‘kind of large forest rat’. Since no marsupial has an indigenous distribution extending east of the Solomons, the name for the rabbit-sized bandicoot evidently survived long enough in Fiji to be applied to larger varieties of placental rat, although cognates are still unknown from Vanuatu, the most likely staging area for the settlement of Fiji and western Polynesia.

4. Savosavo, a non-AN language spoken on Savo island between Florida and Guadalcanal, has *kandora* ‘opossum’, but this clearly has been borrowed from neighboring AN languages, as an identical form is found in Nggela of Florida Island, and in many of the languages of Guadalcanal (Tryon and Hackman 1983:188).

that these words would be borrowed in the same shape into different recipient languages. On the contrary, the normal expectation is that loanwords would be adapted to different phonologies, and that it would therefore be extremely unlikely for multiple borrowing with subsequent descent to result in reconstructable protoforms. The presence of a novel vowel in \*kandoRa might be taken as an indication of borrowing, but this does not bear close scrutiny. As noted in Blust (1993), in addition to the four vowels \*i, \*u, \*ə, and \*a that were inherited from PAN/PMP (where the schwa is written \*e), PCEMP innovated vowels \*e and \*o, at least in part from inherited lexical material. If this innovation had not already taken place, it is likely that a borrowed word of the shape \*kandoRa would have been naturalized as \*kanduRa.

It cannot be denied that there are irregularities in the sound correspondences for reflexes of \*kandoRa and \*mans(aə)r. While Watubela *kadola* points unambiguously to \*kandoRa, for example, Kesui *udora* apparently does not, Buli has *do* for expected \*\*ado, Manam has *ɔdora* for expected \*\*ɔadora, all languages of the Admiralty islands reflect Proto-Admiralty \*kodaya instead of the expected \*\*kadoya, and Nggela of Florida Island along with many languages of Guadalcanal in the central Solomons has *kandora* for anticipated \*\*kandola (Tryon and Hackman 1983:188). However, these are irregularities that characterize the separate histories of individual languages or clearly demarcated subgroups, not products of borrowing related forms of different shape upon first encountering these novel marsupials (Manam *ɔdora* simply shows a sporadic assimilation of the first vowel to the second, Proto-Admiralty shows a sporadic metathesis of the first two vowels, etc.). Vitu *hadora*, Mussau *aroa*, and Lungga *yandora*, on the other hand, point unambiguously to POC \*kadoRa, a form that shows perfect correspondence in form and meaning to Watubela *kadola*. Similarly, some languages support \*mansar and others \*mansər ‘bandicoot’ (a few geographically contiguous languages, as Ngaibor, Ujir, and Kei in eastern Indonesia, point instead to \*mənsar), and most reflexes in Oceanic languages point to initial \*m<sup>w</sup> rather than \*m, a widespread phenomenon in Oceanic languages (Lynch 2002). However, the variation seen in \*mansar ~ \*mansər is part of a much larger pattern of doubling found in many AN languages, and provides no clear evidence of early borrowing from an unattested Papuan source.

Given these observations, it is somewhat disconcerting to see anyone propose such an alternative to the more straightforward hypothesis that \*kandoRa and \*mans(aə)r were innovated by AN speakers upon first encountering marsupials. While it might be argued that the latter hypothesis begs the question of how new lexical items originate, the origin of neologisms is a general problem for which historical linguistics as a whole has never had a fully satisfactory explanation. By contrast, the D&G proposal is a classic example of an ad hoc hypothesis.<sup>5</sup>

5. As noted in Blust (1993:251), Nothofer (1992) has suggested that Popalia (Tukangbesi Islands) *kadola* ‘chicken’, and Malay *buron kedera* ‘Pallas’ mynah: *Agrospar sturnius*’ are cognate with the forms in (1). Given the fundamentally different meanings of the Popalia and Malay forms, both from one another and from \*kandoRa ‘cuscus’, this is a surprising claim. Moreover, it is a claim that commits the fundamental error of ignoring the sound correspondences, since Popalia generally reflects \*R as *h* (occasionally zero, but never *l*): \*ma-wiRi > *mohi* ‘left side’, \*daRaq > *raha* ‘blood’, \*kaRat > *kaha* ‘bite’, \*kaRaw > *kaho* ‘scratch’, \*ma-beRat > *mo-boha* ‘heavy’, \*ma-iRaq > *meha* ‘red’ (cf. <http://language.psy.auckland.ac.nz/>, The Austronesian Basic Vocabulary Database).

The expression “ad hoc” has been so overused as a general term of academic opprobrium that it may be helpful to remind the reader how it is defined by philosophers of science. Both Popper (1968) and Leplin (1975) provide thorough discussions of this term as it applies to certain controversies in the history of physics, in particular to H. A. Lorentz’s “contraction hypothesis,” a claim that formed part of a view of physical reality that was for a time in competition with Einstein’s special theory of relativity. Leplin is at pains to specify conditions under which a scientific hypothesis can be called ad hoc, and toward that end he formulates a number of principles that would be tedious to reproduce here. However, in the course of his discussion he quotes the following passage from Popper: “An example of an unsatisfactory auxiliary hypothesis would be the contraction hypothesis of Fitzgerald and Lorentz that had no falsifiable consequences but merely served to restore the agreement between theory and experiment—mainly the findings of Michelson and Morley” (Leplin 1975:313).

Abstracting away from the specific example at issue, Popper’s point was that a hypothesis is part of science only if it can be falsified (statements that are not falsifiable are part of “metaphysics”). Moreover, a falsifiable hypothesis is justified only if it serves to derive superficially dissimilar sets of observations as expected consequences of a theoretical entity that itself cannot be observed. A classic example is Newtonian gravity: Galileo’s laws of motion for falling bodies on the earth and Kepler’s laws governing the elliptical orbits of planets around the sun were known and accepted well before Newton, but they had been treated as unrelated generalizations about physical reality. A significant part of the success of Newton’s gravitation theory is that it accounted for both Galileo’s laws and Kepler’s laws as the expected consequences of a wider generalization. An ad hoc hypothesis, on the other hand, is one that serves no purpose other than to allow the theorist to avoid confronting an inconvenient set of observations.

The explanation that D&G propose for the distribution of cognate marsupial terms in AN languages fits this definition of an ad hoc hypothesis closely: the cognate sets for marsupial mammals are inconvenient for anyone who espouses the subgrouping represented in figure 2, since there would then be no obvious way to explain why these terms permit unambiguous protoforms to be reconstructed. To avoid this inconvenience, an unattested Papuan lingua franca is invented, and the nonconforming evidence is swept under the theoretical rug, to be quickly forgotten. But this is not the way historical linguistics normally works. Rather, the usual assumption is that exclusively shared innovations reflect a period of exclusively shared history unless it can be shown that they are likely products of convergence, or are demonstrable loans acquired independently in the languages that contain them. Reflexes of \*kandoRa and \*mans(aə)ɾ thus provide straightforward support for an inference that at least many of the languages of eastern Indonesia subgroup immediately with Oceanic. According to Darlington (1980:374) “*Phalanger* (cuscuses) sets both western and eastern limits of marsupials in the Indo-Australian Archipelago: the genus is primarily New Guinean (2 species of it extend to the upper part of the Cape York Peninsula of Australia), and it occurs west to Timor and Wetar and to Celebes, and east to the remotest Solomons.” What this means is that the absence of related terms in CEMP languages spoken to the west of Timor is a consequence of the (current) faunal distribution itself: without referents, there can be no terms of reference. But the inclusion of these

other languages of eastern Indonesia within CEMP follows from other lines of evidence, as will be shown below. Notably, although Sulawesi has two distinct, endemic species of phalanger, none of the terms for these animals in the languages of Sulawesi is related to \*kandoRa, an observation that is consistent with the hypothesis that AN speakers encountered the cuscus independently in Sulawesi and in the Lesser Sundas-Moluccas, and innovated distinct names for it in these two regions.<sup>6</sup>

In short, there is no well-argued alternative to the hypothesis that \*kandoRa ‘cuscus, phalanger’ and \*mans(aə)r ‘bandicoot’ were lexical innovations in a single language community that was immediately ancestral to most or all of the AN languages of eastern Indonesia and of the Oceanic subgroup.

**3.2 PCEMP \*kera(ŋ) ‘HAWKSBILL TURTLE’.** A third faunal term of considerable interest is \*keRa(ŋ) ‘hawksbill turtle’. Although \*peŋu ‘green turtle’ and \*qaCipa ‘softshelled freshwater turtle’ can be assigned to PAN, a word for ‘hawksbill turtle’ has not been reconstructed for PAN or PMP, nor are distinct words for this type of turtle given in most dictionaries of languages in insular Southeast Asia (Blust 2002:121). The known evidence suggests that PAN/PMP \*peŋu may have referred to both the green turtle and the hawksbill, and that these referents may not have been lexically distinguished until AN speakers reached eastern Indonesia. The evidence from Blust (1993) is recapitulated in (3), with additions.

(3) PCEMP \*keRa(ŋ) ‘hawksbill turtle’

CMP:

Manggarai	héra ‘sea turtle’
Ngadha	kera ‘large turtle’
Lio	kera ‘tortoise, land turtle’
Sika	era ‘turtle’
Hawu	ea ‘turtle, turtle shell’
Roti	kea ‘turtle’
Leti	kema ‘marine turtle w. prized shell: <i>Chelonia imbricata</i> ’
Fordata	eran ‘turtle with prized shell’
Kei	keran ‘turtle with prized shell’

OC:

Motu	era ‘turtle; shell turtle’
Kapingamarangi	gee ‘hawksbill turtle: <i>Eretmochelys imbricata</i> ’
Nukuoro	gea ‘hawksbill turtle’
Rennellese	kea ‘small turtles’
Nanumea	kea ‘a small turtle’
Hawaiian	?ea ‘hawksbill turtle; the shell of this turtle’

Blust (1993) reconstructed disjuncts \*keRa and \*keRa(ŋ) for this comparison on the grounds that Manggarai retains final nasals, and hence *héra* ‘sea turtle’ could not reflect

6. Cf. Proto-Sangiric \*kusay (Sneddon 1984:89), Sangir *kumbanaj*, *kuse*, Proto-Minahasan \*kuse, Bolaang Mongondow *kutoy*, *kokilobon*, Bare’e (Pamona) *kuse*, *tumapo* (Pu’umboto dialect), *wajo*, *tajali* (Napu dialect), Tae’ *kuse*, Proto-Bungku-Tolaki \*kuhe (Mead 1998:444), Muna *kampaghi*, *kuhe*, *loghu*, Wolio *kuse*, Makasarese *memu*—all referring to types of cuscus and supporting Proto-Celebic \*kusay, with apparent borrowing into the Sangiric, Minahasan, and Gorontaloic languages, which arrived at a later date. The bandicoot is unknown in Sulawesi.

\*keRa(ŋ), but could be compared with forms such as Hawaiian *ʔea*, which are ambiguous for the presence or absence of a final consonant. However, final \*ŋ did disappear in a few Manggarai forms, as with \*kedenj > *hesé* ‘stand’, and \*qudarj > *k-usé*, next to the apparently borrowed *kuranj* ‘shrimp’ (Blust 2008:87). Since all other reflexes that were assigned to \*\*keRa are ambiguous for a final nasal or zero, we can settle on \*keRa(ŋ) as the only justified shape of this reconstruction.

Unlike \*peŋu, which is richly attested over a wide area, \*keRa(ŋ) has a surprisingly light and scattered distribution (and consequently the recognition of this form was decades behind that of \*peŋu). Despite thorough searching of available published sources, reflexes of \*keRa(ŋ) have so far been found only in a small number of languages in the Lesser Sundas and southern Moluccas of eastern Indonesia, and in six Oceanic languages. Nonetheless, the Oceanic languages are widely separated (Motu and Polynesian). Even within the Polynesian subgroup the distribution is surprisingly sparse and scattered, including Rennell-Bellona, which is a primary branch of Nuclear Polynesian, three Ellicean Outliers (Kapingamarangi, Nukuoro, and Nanumea), and a single Eastern Polynesian language (Marck 2000:3). Despite this relatively light attestation, both the form and its associated meaning can confidently be reconstructed.

Although \*keRa(ŋ) is not a replacement innovation (since a term for ‘hawksbill turtle’ has not been reconstructed for PAN or PMP), its distribution clearly has important implications for subgrouping. First, like \*kandoRa and \*mans(aə)r, it is found in eastern Indonesia and the Pacific, but is unknown elsewhere. Second, unlike the marsupial terms, which are lacking in the western Lesser Sundas, reflexes of \*keRa(ŋ) are found in the languages of Flores, Hawu, and Roti, thus linking these languages with those of the Moluccas and the Oceanic group. Third, like \*kandoRa, \*keRa(ŋ) has an innovative vowel, a fact that argues against it being assignable to any higher-level protolanguage. Nothofer (1992) suggested that Padoe *kéha* and Kada Toraja *kéra* ‘shell’ might be assignable to \*keRa(ŋ), but if these forms are related, they almost certainly are loanwords, since (1) the Toraja languages are spoken in interior Sulawesi, and their speakers would have no knowledge of the hawksbill turtle apart, perhaps, from its valued shell, which would have been acquired in trade, and (2) as already noted, the vowel \*e (not \*ə) has not been reconstructed for PAN or PMP. The argument from borrowing might, of course, be turned back against the languages in (3): if possible reflexes of \*keRa(ŋ) in Toraja languages of central Sulawesi are best explained as loanwords, then what would prevent the same explanation being invoked to account for the distribution in (3)? The fundamental difference, of course, is that the shell of the hawksbill turtle is prized, and it is this that is traded, not the turtle itself, whereas all reflexes in (3) refer to the hawksbill turtle, and if they refer to its shell, it is part of a more general gloss that also includes the animal from which the shell is obtained. In conclusion, \*keRa(ŋ) ‘hawksbill turtle’ is a valuable etymology that cannot simply be ignored, as D&G have done.

These three etymologies alone—\*kandoRa ‘cuscus’, \*mans(aə)r ‘bandicoot’, and \*kera(ŋ) ‘hawksbill turtle’—are sufficient in themselves to make a strong case that most or all of the languages of eastern Indonesia subgroup immediately with Oceanic, as opposed to, for example, various of the languages of Sulawesi. If this hypothesis is rejected, it becomes very difficult to explain the distribution of these terms. Borrowing from a

hypothesized Papuan lingua franca is little more than an artifice for dismissing clearly innovative terms for marsupial mammals that are shared by a number of the languages of eastern Indonesia with Proto-Oceanic, and claims that cognates of PCEMP \*keRa(nj) are found in some of the languages of central Sulawesi disappear on closer inspection.

**3.3 OTHER LEXICAL INNOVATIONS.** Although a confident subgrouping hypothesis can be built on the preceding lexical innovations alone, other innovations in the lexicon point to the same conclusion. The more important of these that have come to light so far include the following; for convenience, even though they are in dispute, the labels “Central Malayo-Polynesian” (CMP), and Central-Eastern Malayo-Polynesian (CEMP) will be used during the course of justifying these proposed subgroups.

**3.3.1 \*kanzupay ‘rat’.** PMP \*labaw can be securely reconstructed in the meaning ‘rat’. In some languages this form is reflected alone, and in others with apparent prefixes \*ba- or \*ka-. A small number of possible reflexes are found both in eastern Indonesia and in Oceanic languages. In addition to these, however, reflexes of another term, \*kanzupay ‘rat’, appear in a number of the languages of the central and northern Moluccas, as (CMP) Gah *karufei*, Watubela *arofa* (Wallace 1962), Waru *karufa* (Stokhof 1982:65), (SHWNG) Mysol *keluf* (Wallace 1962), Buli *luf* ‘rat’ (Maan 1940), and in Oceanic languages reaching from the Bismarck archipelago and southeast New Guinea to Fiji (Milke 1968:152). This term is unknown elsewhere in the Austronesian language family.

**3.3.2 \*liqəR ‘voice’.** (CMP) Tetun *lia-n* ‘voice, the sound of the voice, the sound made by anything, noise’, Leti, Wetan *lira* ‘sound, voice, word’, Buru *lie* ‘words’, *lie-n* ‘voice, words’ (Devin 1989), Proto-Ambon \*liə ‘voice’ (Stresemann 1927:22), (SHWNG) Buli *lio* (expected \*\*liā) ‘voice, sound’, (OC) Lau *leo* ‘speech; but only in names in genealogies’, *lio* ‘voice ... only in compounds’, *lio sī* ‘hoarse’, Rotuman *lio* ‘voice’, Wayan *lio* ‘voice, of any living thing’ (Pawley and Sayaba 2003), Proto-Polynesian \*leʔo ‘voice’ (Walsh and Biggs 1966:47). This is possibly a reflex of PMP \*liqəR ‘neck’ with semantic shift, but this seems unlikely. This form is unknown outside CMP, EMP, and Oceanic, and appears to have replaced PMP \*huni ‘chirp, tweet, crow, produce sound (of birds)’ (but also Cebuano *húni* ‘song, musical sounds’, Kelabit *unih* ‘sound’, Old Javanese *uni* ‘sound, voice’).

**3.3.3 \*malo ‘loincloth’.** Although its known distribution includes only languages in the Philippines and Borneo, it is possible that \*bahaR ‘loincloth’ was found in PMP. If so, it was replaced by reflexes of \*malo in eastern Indonesia and the Oceanic group: (CMP) Selaru *mal-ke*, *mala-re*, Buru *maru* ‘loincloth’, (SHWNG) Buli *māl* ‘pounded tree bark; clothing of same’, Numfor *mār* ‘loincloth (originally of loincloths made of pounded tree bark)’, POC \*malo ‘paper mulberry: *Brousonettia papyrifera*; loincloth’. While the Buli and Numfor forms can be compared directly with POC \*malo, and the Selaru forms may point to a comparable stem *mal* (rather than *mala*), Buru *maru* shows *r* for expected *l* and *u* for expected *o*. These discrepancies and the rarity of cognates in island Southeast Asia make this comparison less than perfect, but it is also clear that the resemblances in question cannot easily be attributed to chance or borrowing.<sup>7</sup>

**3.3.4 \*malip ‘laugh’.** (CMP) Hawu *mari*, Atoni *mani*, Galoli *rar-mali*, Leti *mali*, Yamdena *malip*, Fordata *malit*, Ngaibor *mel*, Geser *malif*, Sekar *manif*, Koiwai *marif*, Sub-Ambon \**ma-lipe* (Stresemann 1927:52) ‘laugh’, (SHWNG) Gimán *mlif*, Minyaifuin *mnif* ‘laugh’, (OC) Seimat *mal* ‘laugh, smile’, *malimal* ‘laughing, smiling’, Proto-Chuukic \**mali*, \**mmali* ‘smile’ (Bender et al. 2003), Fijian *mali* ‘to smile’, *malimali* ‘smiling’, Wayan *mali* ‘to laugh, smile, grin’, Tongan *malimali*, Niue *mamali* ‘to smile’. Although Kédang and the languages of Flores retain a reflex of PMP \**tawa* ‘to laugh’, reflexes of \**malip* are widely scattered throughout the rest of eastern Indonesia and Oceanic, and are clearly innovative.

**3.3.5 \*tosi ‘make a mark, scratch a line’.** (CMP) Atoni *tusi* ‘draw a picture; write’, (SHWNG) Serui-Laut *toi*, *toi-na* ‘to write’ (Anceaux 1961:28), (OC) Arosi *osi* ‘to cut, score’, Wayan *tosi* ‘make a mark on a surface, mark, draw, etch’, Tongan *tohi* ‘make marks or designs on (e.g., tapa cloth); write, write about’, Samoan *tosi* ‘score, scratch’, *tusi* ‘draw, write’, Rennellese *tosi* ‘to write, draw, outline a tattooing pattern’, *tusi* ‘to mark, as the skin with turmeric’. Since writing was unknown in eastern Indonesia and the Pacific prior to European contact, this term presumably meant something like ‘scratch a line, draw’. Most of the comparative evidence for this term points to \**tosi*, but a subset of forms instead indicates \**tusi*.

**3.3.6 \*tambu ‘taboo, forbidden’.** (CMP) Yamdena *tambu* ‘restrain, prevent, forbid’, Fordata *tabu* ‘forbid, prevent’, POC \**tabu* ‘forbid, prohibit’. It is clear from a number of reflexes in Nuclear Micronesian languages that PMP \**palihi* ‘taboo, forbidden’ persisted in POC, and these two terms presumably were distinguished semantically in some way that has yet to be clarified. In any case, known reflexes of \**tambu* are restricted to the southern Moluccas and Oceanic.

**3.3.7 \*wajka ‘canoe, boat’.** (CMP) Komodo, Manggarai, Rembong *wajka* ‘boat, canoe’, Buru *waga* ‘canoe, boat, ship’, Proto-Ambon \**waga* ‘canoe’, (SHWNG) Mor *waʔa*, Dusner *wak* ‘canoe’, Numfor *wa(i)* ‘single or double outrigger canoe. Used only of canoes with dugout hulls’, Waropen *gha* ‘boat, canoe’, (OC) Wuvulu, Seimat *wa* ‘canoe’, Tangga *waj* ‘outrigger canoe of dugout type’, Label *waga*, Duke of York *anga* ‘canoe’, Gedaged *wag* ‘large canoe that goes out on the high seas, has one or two masts and a large platform; ship, boat’, Gitua *wajga*, Dobu *waga* ‘canoe’, Bwaidoga *waga*, *waka* ‘a canoe of any kind, but esp. a large, boarded-up canoe’, Roviana *vaka* ‘a vessel—used distinctively of a nonnative craft’, *tie vaka* ‘white man’, Ghari *wajga*, Tanimbili *n-ongo* ‘canoe’, Kiribati *wa* ‘any vehicle, any means of conveyance; canoe, boat, bicycle, cart’, Marshall-ese *wa* ‘canoe; ship; boat; vehicle’, Chuukese *waa* ‘boat, ship, canoe, vessel, conveyance, automobile, airplane, vehicle’, Puluwat *waa* ‘canoe, vehicle of any kind, bicycle, container, people in a canoe; possessive classifier for vehicles’, Mota *aka*, Hiw *wakə*, Lehali *n-ok*, Jawe *waj*, Rotuma *vaka* ‘canoe’, Fijian *wajga* ‘a sailing canoe, of various kinds’, Proto-Polynesian \**waka* ‘canoe, boat’.

Based on the comparison Javanese *wajkanj* ‘ship’, Ngaju Dayak *wajkanj* ‘Chinese junk’, Fijian *wajga*, Tongan, Futunan *vaka*, Samoan *vaʔa* ‘canoe’, Dempwolff (1938)

7. The attribution of a more restricted form of this comparison to a higher-level protolanguage in Blust (1970:133) has clearly been superseded by subsequent research.

reconstructed \**wanʔkaŋ* ‘water vessel, ship, canoe’. This long accepted etymology actually is highly problematic. First, reflexes with a final nasal are known in only five languages, and all of these refer specifically to Chinese junks: Ngaju Dayak *wanʔkaŋ* ‘Chinese junk’, *tanah wanʔkaŋ* ‘China’, Malay *wanʔkaŋ* ‘Chinese ocean-going junk’, *limau wanʔkaŋ* ‘imported Chinese oranges’, Acehnese, Javanese, Makasarese *wanʔkaŋ* ‘Chinese junk (boat)’. Second, neither the Malay nor the Javanese words can regularly reflect \**wanʔkaŋ*, since \**w-* disappeared in Malay and \**wa-* became Javanese *o*. Moreover, the word in Ngaju Dayak, which is spoken in the interior river systems of Southeast Borneo, clearly is a loan from the coastal Banjarese Malay. It appears likely, then, that *wanʔkaŋ* is a loan from some southern variety of Chinese that was borrowed into Malay and then spread from it to other languages of western Indonesia. In apparent contradiction to this conclusion are reflexes in CMP, SHWNG, and especially OC languages, where the probability of early borrowing from Chinese is all but nonexistent. Since most OC languages have lost original final consonants, many OC reflexes could derive from \**wanʔkaŋ*. In those languages that normally preserve the final consonant, however, no final consonant is found in this word (Dobuan, Roviana), thus supporting \**wanʔka*, not \**wanʔkaŋ*. It appears simplest, then, to conclude that \**wanʔka* was an innovation that appeared in northeast Indonesia prior to the separation of the CMP and EMP language groups from one another, but after the separation of PCEMP from all other AN languages, and that its resemblance to the loanword *wanʔkaŋ* is a striking coincidence.

**3.3.8 Other innovations.** This is only a small sample from a collection of 33 proposed innovative etymologies in Blust (1993), which does not include \**kanzupay*, \**tos*, or \**wanʔka*. Moreover, some thirty other proposed lexical innovations supporting CEMP appear in Blust (1995), which has been available upon request for some years, and has in fact been extensively used in other major studies of the lexicon of early AN protolanguages, such as Ross, Pawley, and Osmond (1998, 2003). It is important to note that the Austronesian Comparative Dictionary (ACD) has been compiled through a systematic comparison of over 150 languages, and the probability is therefore rather small that reconstructions that are assigned to PCEMP in that work will turn out to be invalid either because they are based on erroneous cognate decisions, or because they are found in other AN languages. In addition, it should be stressed that although it is over 2,000 pages in computer printout, the ACD is still only about 25 percent complete. Space does not permit the full citation of supporting evidence, but the following PCEMP reconstructions can be added to those already cited.<sup>8</sup>

(4) * <i>abat</i> ‘wound’	* <i>buu</i> ‘to blow; conch shell trumpet’
* <i>au</i> ‘dew’	* <i>buu</i> ‘corner, edge, point’
* <i>ba</i> ‘go’	* <i>ərit</i> ‘scratch, scrape’
* <i>balan</i> ‘side, part’	* <i>i</i> ‘numeral prefix’
* <i>belen</i> ‘to swallow’	* <i>ima</i> ‘pandanus w. leaves useful for plaiting’
* <i>bina</i> ‘shellfish sp.’	
* <i>birinj</i> ‘to stone, throw a stone’	* <i>isi</i> ‘peel, strip off’

8. Note that protolanguages in the ACD are coded as (1) PAN, (2) PMP, (3) PWMP, (3a) PPH, (4) PCEMP, (5) PCMP, (6) PEMP, (7) PSHWNG, and (8) POC. Reconstructions in which (4) precedes the headword are PCEMP; those in which (4) is embedded in a larger comparison represent the PCEMP continuations of higher-level reconstructions. Only the former are counted here.

*bisik ‘to peel’	*matay ‘money, payment, medium of common exchange’
*bitak ‘mud’	*pali ‘side, half’
*bitu ‘sword grass: <i>Imperata cylindrica</i> ’	*paniŋ ‘bait’
*bua ‘only, just’	*papaR ‘cheek’
*bubu ‘sing; song’	*paRa- ‘reciprocal’
*budeq ‘sponge’	*qaue ‘exclamation of joy or sorrow’
*buqal ‘levered up, uprooted’	*quRi ‘a fruit tree: <i>Spondias dulcis</i> ’
*butu ‘group, crowd, flock, school, bunch, cluster’	*wakir ‘k.o. root’
	*wari ‘sing; song’

A few other comparisons are suggestive, but violate the strict distributional limitations needed to justify a CEMP group. The most striking of these is \*qumun ‘earth oven’. Although earth ovens have a scattered distribution globally, the use of earth ovens in the AN world appears to be confined to eastern Indonesia and the Pacific. Chowning (1991:55) notes that “the term ‘earth oven’ is a misnomer in many parts of Melanesia, including Lakalai, Kove, and Sengseng, where the whole process is completely above ground.” Nonetheless, whether it is excavated or built upon the surface of the ground, the earth oven is clearly distinct from other cooking devices, and the association of this structure with a reflex of \*qumun in several of the languages of the central and southern Moluccas and many Oceanic languages is noteworthy. What appears to vitiate the use of this comparison as evidence for a CEMP subgroup is the appearance of Palauan *chum* ‘hole in the ground for baking’, *məŋ-úm* ‘bake food in a hole in the ground’. As noted in Blust (1995), it is tempting to consider Palauan *chum* as a loan from an Oceanic language since (1) it is the only reflex of \*qumun yet reported from any non-CEMP language, and (2) it appears to show an irregular loss of the final consonant (expected \*\*chuml). However, *məŋ-úml* ‘start up, kindle (fire)’ may also reflect the same etymon without loss of the final, and other features of this comparison suggest that it has been in Palauan for a considerable period of time, so if it is a loanword it is one that entered the language very early.

**3.4 OTHER EVIDENCE FOR CEMP IN BLUST (1993).** After postponing, and therefore effectively dismissing, the lexical evidence for CEMP, D&G summarize the remaining evidence presented for CEMP in Blust (1993) in their tables 15 and 23. For convenience of reference, their table 23 is reproduced here as table 1, with the addition of “Regular phonological changes,” which they include in their table 15, but not in the later table.

**3.4.1 Cluster reduction, 1.** Proto-Austronesian had only one type of consonant cluster, namely a sequence of two (generally heterorganic) consonants found in monomorphemic reduplicated monosyllables such as \*tuktuk ‘to knock, pound, beat’. PMP preserved these clusters, and to them added (generally homorganic) prenasalization of medial obstruents, as in \*tumbuq ‘to grow’ or \*punti ‘banana’. Blust (1993) noted that reduction of medial clusters is found in a few languages of the Philippines and western Indonesia, but appears to be universal in CEMP, and therefore is most simply explained as the product of a single historical change. D&G (2008:134) hold that cluster reduction is “an expected feature of any language that approximates a CV syllable structure.” The problem with this objection is that many of the languages of eastern Indonesia and some

Oceanic languages have preserved word-final consonants, and so are not languages with a CV syllable structure. This is true of most of the languages of Flores, Timor, the southern, central, and northern Moluccas, and of various parts of western and southern Melanesia, where POC final consonants are retained, sometimes with and sometimes without an echo vowel, or a supporting vowel *-a*.

The second objection that D&G raise to medial cluster reduction as evidence for CEMP depends upon the analysis of a single language in an unpublished conference paper. Grimes (1991) claims that evidence for medial clusters is preserved in Buru, since these were not always reduced by loss of the first member. D&G (2008:135) give three examples of cluster reduction in which the pattern of reduction reportedly is  $C_1V_1C_2 C_1V_1C_2 > C_1V_1C_2V_1C_2$  rather than  $C_1V_1C_2 C_1V_1C_2 > C_1V_1C_1V_1C_2$ . These are \*basbas ‘cut’ > *fasa* ‘cut, decide’, \*kepkep ‘hold’ > *gepe* ‘hold tightly in hand’, and \*kiskis ‘scrape, grate’ > *kisi* ‘stroke with fingertip’. The first of these comparisons evidently depends upon the reconstruction \*basbas ‘trim, prune, as branches, weeds, etc.’ (Blust 1995). The only dictionary that is available for Buru is Devin (1989), which gives *fasa-h* (vt) ‘dissect, ford, cut through, decide’, *fasa-k* (vt) ‘cut, prepare’ (with morpheme boundaries added). In the absence of other evidence, it is impossible to decide whether these forms actually reflect \*basbas, or have some other source. The comparison is hardly compelling as given, and further examples of a similar type and better quality would be needed to make a convincing case. More serious methodological objections apply to the historical

**TABLE 1. THE CEMP INNOVATIONS RECONSIDERED  
(FROM DONOHUE AND GRIMES 2008)**

	Defines CMP?
<b>Regular phonological innovations</b>	
1. Cluster reduction, 1: $C_1V_1C_2 C_1V_1C_2 > C_1V_1C_1V_1C_2$	no
2. Cluster reduction, 2: $CVN_iC_jVC > CVN_jC_jVC$	no
<b>Irregular phonological innovations</b>	
3. PMP *uliq > PCEMP *oliq ‘return’	no
4. PMP *i-sai > PCEMP *i-sei ‘who’	no
5. PMP *ma-qitəm > PCEMP *ma-qetəm ‘black’	no
6. PMP *maRi > PCEMP *mai ‘come’	no
7. PMP *tudan > PCEMP *todan ‘sit’	(yes)
8. PMP *inum > PCEMP *unum ‘drink’	no
<b>Morphosyntactic innovations</b>	
9. Prefixal/proclitic agreement on verb	no
10. Alienable/inalienable possession distinction	no
11. Frozen morphology	
a. PMP *həpat > PCEMP *həpat, *pat, *pati ‘four’	no
b. PMP *ma-huab > PCEMP *mawab ‘yawn’	no
c. PMP *ma-hiaq > PCEMP *mayaq ‘shy’	no
<b>Semantic innovations</b>	
12. PMP *t-ina ‘mother’ > PCEMP *t-ina ‘big’	no
13. PMP *m-udehi ‘behind’ > PCEMP *mudi ‘back (of body)’	no
14. PMP *ma-qitəm ‘black’ > PCEMP *ma-qetəm ‘dirty’	no
15. PMP *tuqəla(ŋ) > PCEMP *zuRi ‘bone’	no
16. PMP *buhək > PCEMP *daun ni qulu ‘head hair’	no
17. PMP *daləm ‘inside’ > PCEMP *daləm ‘mind, feelings’	no

derivation \*kepkep ‘grasp, hold’ > *gepe* (Devin gives *gepeh* [vt] ‘clamp’). PMP \*k normally became Buru *k*, as in PMP \*kali > *kali-k* (vt) ‘dig’, \*kasaw > *kasa ~ kaso* ‘roof pole’, \*keden > *kere* ‘stand’, \*kita > *kita* ‘we (incl.)’, \*kutu > *koto* ‘louse, flea, tick’, or \*kuden > *kuren* ‘large earthenware jar’. D&G (2008:135) assert that \*k > g is “semi-regular,” but give no support for this statement. Devin (1989) contains eleven pages of Buru forms that begin with *g-*, and I am unable to trace any of these to an etymon with \*k-. Only slightly less systematic searching of the history of Buru medial *g* shows similar results. Moreover, there are no plausible etymologies that show the type of cluster reduction claimed, which is in fact unattested in any known AN language. D&G (2008:134, fn. 25) try to shore up this clearly shaky proposal by offering a back-up explanation, namely that Buru *gepe* may reflect “an unreduplicated monosyllable \*kep, and an epenthetic vowel echoing the quality of the vowel in the preceding syllable. This interpretation is equally valid; regardless of which is chosen, the cluster reduction pattern described by Blust does not account for the Buru facts.” Apart from the fact that unreduplicated monosyllables cannot be reconstructed unless they are onomatopoeic (Blust 1988:1–2), the hypothesis of an obviously sporadic echo vowel is simply another example of the rather desperate “explanations” that D&G resort to repeatedly throughout this paper in an effort to find alternatives to the CMP and CEMP hypotheses.<sup>9</sup> The last example, which Devin (1989) gives as *kisi-h* (vt) ‘remove (as eye matter), poke out, pry out, debone’, is also problematic. First, the semantic agreement with \*kiskis is poor. Second, it can be compared with Ngadha *kisi* ‘peel or pare off, strip or draw off with the finger’, and Fordata *kisi* ‘pinch, hold between thumb and forefinger’, and so need not go back to \*kiskis at all. This leaves little choice but to conclude that D&G have relied primarily on guesswork rather than on sound comparative procedures in claiming that Buru sometimes retains the first consonant of the cluster in reflexes of PMP reduplicated monosyllables.

The fact remains that although a comparatively small number of AN languages outside eastern Indonesia and the Oceanic group have reduced medial clusters in reduplicated monosyllables, all of the 600 or more languages that have been assigned to CEMP reflect this innovation, and the simplest hypothesis to explain this distribution is one that posits a single change in a language immediately ancestral to them and no others.

**3.4.2 Cluster reduction, 2.** D&G (2008:135–36) also treat place assimilation for nasals in cases such as PMP \*demdem > PCEMP \*dendem ‘dark’ as though it was used as subgrouping evidence in Blust (1993). But attention to the original text (1993:246) shows that this is not the case. Rather, it is simply noted there that while the first conso-

9. A compilation of all Buru forms with *g* for which an etymology can be established shows that these have either of two sources: (1) they derive from *g* in Malay loanwords (*gaji-h* < *gaji* ‘pay, hire’, *dagan* < *dagan* ‘foreign’, etc.), or (2) they derive from PMP \*g in words that appear to be native (*gege* < \*gerger ‘tremble’, *pager* < \*pager ‘fence’). The only cases in which it might be argued that Buru *g* reflects PMP \*k are *gaan* ‘a tree: *Inocarpus fagifera*’, *gae-h* (vt) ‘hook’, and *gilak* ‘shine’. The first of these can be compared with Taosug *kayam* (Madulid 2001:369), the second with PMP \*kawil ‘fish hook’ or \*kawit ‘hook’, and the third with PMP \*kilat ‘lightning’. The problem with such speculations is that, for language-universal aerodynamic reasons, the voicing distinction for \*k and \*g is known to be inherently unstable (Blust 1996). As a result, one could as easily compare Buru *gaan* with Iban *gayam* ‘a tree: *Inocarpus edulis*’, or Buru *gilak* with Cebuano *gilak* ‘glitter, sparkle’. As for Buru *gae-h*, this comparison clearly loses its initial appeal in view of Buru *kawi-k* (vt) ‘hook’, *kawil* ‘fishhook’.

nant of a medial cluster normally dropped, it was retained with place assimilation if it was a nasal followed by an obstruent. No further claim was made, and therefore no further claim needs to be challenged.

**3.4.3 PMP \*uliq > PCEMP \*oliq ‘return’.** Blust (1993) cited this change for three reasons: (1) it appears to be lexically specific, (2) it introduced a vowel that was not part of the phoneme inventory of PMP, and (3) it is found in languages of the Lesser Sundas (Manggarai, Hawu, Lamaholot), the southern and central Moluccas (Selaru, Yamdena, Buru), and in a number of the languages of the Solomon Islands, including Bugotu, Nggela, Kwaio, Sa’a, Arosi, and ’Are’are. The gloss given to PMP \*uliq in Blust (1995) is ‘return home; return something; restore, repair; repeat; motion to and fro’. D&G wish to dismiss this innovation on the grounds that forms with unlowered \*u are found in some of the languages of the Lesser Sundas, as Palu’e *phuli* ‘return (parallel speech form)’. The entry in Blust (1995), which D&G do not cite, in fact shows that whereas reflexes of \*oliq are dominant in CEMP languages, reflexes of \*uliq also occur, as in the only known Polynesian reflex, Rarotongan *uri* ‘turn around or turn about; translate, as one language into another’.

These facts are not in question. What is in question is D&G’s clear departure from the logical basis of the Comparative Method in subgrouping languages, as described in section 2. For the change \*uliq > \*oliq to carry weight as subgrouping evidence, all that matters is that convergence can be ruled out as a plausible explanation for the shared innovation, and that this innovation is exclusively shared by the languages in the proposed subgroup. As will be seen in proceeding to other examples, this point needs to be emphasized, since a number of the proposed innovations defining PCEMP are dismissed on similar grounds.

Attention to the entry for \*uliq in Blust (1995) shows a single form \*uliq for PMP, coded (2), but coexisting variants \*oliq, \*uliq for PCEMP, coded (4). This is because there is no convincing evidence for the variant \*oliq in WMP languages, but both \*uliq and \*oliq are widely reflected in CEMP languages. The force of this innovation as subgrouping evidence stems, of course, from its being lexically specific, since the probability that a sporadic lowering of \*u would affect the same morpheme independently in languages of eastern Indonesia and the Solomon Islands is vanishingly small. Real damage to the claim expressed in this comparison would follow only if it were demonstrated that some WMP languages show a similar sporadic change in reflexes of PMP \*uliq. It is possible that Muna *doli* ‘turn around, turn one’s head’ is connected, and also shows sporadic lowering of \*u, but this is by no means certain. Other cases that at first appear to show such lowering fall apart under close scrutiny. Mills (1975:2:879) reconstructs Proto-South Sulawesi (PSS) \*ole (?\*ule) ‘again, return’, for example, but this form shows lowering of both vowels, and the lowering of a penultimate high vowel is attested in some other forms, as PWMP \*guyan > PSS \*goyan ‘shake, waver’, PMP \*linduŋ > PSS \*lendonŋ ‘eel’, PMP \*qulu > PSS \*olo ‘in front; before; earlier’, and PMP \*tuqtuq > PSS \*toto ‘trim, prune branches’.<sup>10</sup> All in all, then, the lowering of the first vowel in PMP \*uliq appears to be sporadic, not clearly attested in any WMP language, and unlikely to be a product of convergence.

**3.4.4 PMP \*i-sai > PCEMP \*i-sei ‘who?’.** This proposed innovation provides weaker support for CEMP than the preceding, since convergence is more difficult to rule out. Again, D&G insist arbitrarily on excluding changes of the type  $*A > A \sim B$  as subgrouping evidence. Surprisingly, in view of the fact that they themselves recognize this as a “phonetically natural” change, D&G (2008:137) cite Proto-Kaili-Pamona *\*i-sei-ma* ‘who?’ and assert that this is “strong evidence that the irregular development of *\*i-sai > \*isei* was found earlier than would be suggested by the borders of ‘CMP’ languages.” Why this distribution should be cited as evidence for an improbable linguistic subgroup that includes the Kaili-Pamona languages together with many of the languages of the Lesser Sundas and Moluccas rather than as an example of convergence is neither explained nor justified.

**3.4.5 PMP \*ma-qitəm > PCEMP \*ma-qetəm ‘black’.** This comparison raises an important methodological issue. If a sound change is regular and of a common type, it has little or no subgrouping value by itself, since common sound changes are common precisely because they are products of multiple historical events. If a sound change is unique to a particular morpheme, on the other hand, it has great subgrouping value, since it is unlikely that a lexically specific change would target the same morpheme through two or more historically independent events. Sound changes are often portrayed as belonging to either of these two extreme types: completely regular, or lexically unique. However, some sound changes are sporadic but nonunique, and these present a more delicate problem in evaluating the probability that they are products of a single historical event or of multiple historical events.

PMP *\*ma-qitəm > PCEMP \*ma-qetəm* initially appears to present a serious challenge to the CEMP hypothesis. At least two languages in eastern Indonesia reflect *\*ma-qitəm* unambiguously with sporadic lowering of *\*i*: Watubela *ma-ketan* and Koiwai *ma-etan* ‘black’. Other languages of eastern Indonesia are ambiguous for a development *\*ma-qitəm > maitəm > metəm*, vs. *\*ma-qitəm > maetəm > metəm*. The irregular change in at least Watubela and Koiwai agrees with the lexically specific lowering of *\*i* in Oceanic reflexes such as Peterara, Wailengi, Tasmate *maeto*, or Tongariki, Namakir *ma?et* ‘black’, and so has strong *prima facie* subgrouping value. However, in both eastern Indonesia and in Oceanic, reflexes of PMP *\*ma-qitəm* also occur without lowering: Manggarai, Kédang *miteŋ*, Palu’e, Ende *mite*, Nasawa/Baetoro *maito* ‘black’. Reflexes of PMP *\*ma-qitəm* thus follow the generalized form  $*A > A \sim B$ . As seen above, where they encounter this form with *\*uliq > uliq ~ oliq* D&G reject the claim that an innovation has occurred. Unexpectedly, with PMP *\*ma-qitəm > \*ma-qitəm ~ \*ma-qetəm* they opt instead for a hypothesis of innovation, but one that includes some of the languages of Sulawesi together with others of eastern Indonesia.

Following Nothofer (1992), D&G point out that a similar sporadic lowering of *\*i* in reflexes of *\*ma-qitəm* is found in several Kaili-Pamona, Wotu-Wolio, and Bungku-Tolaki languages of central and southeast Sulawesi (while others reflect *\*ma-qitəm* with-

10. The lowering of high vowels before final *\*q* appears to be regular: PMP *\*buluq > PSS \*bulo* ‘bamboo’, PMP *\*bunuq > PSS \*buno* ‘strike, kill’, PMP *\*puluq > PSS \*pulo* ‘tens’, etc. It is possible that this influenced the lowering of *\*u* in PSS *\*ole*, but this would not be regular, as South Sulawesi languages lack vowel harmony.

out lowering), and that it therefore “appears highly likely that the variant \*ma-(q)eta was present in Proto-Celebic.” Among these languages, only Wolio, Uma, and Pamona (Bare’e) are represented by dictionaries or extensive wordlists that provide enough evidence for detailed understanding of the historical phonologies. Since a unique lowering of \*i in the same morpheme is unlikely to be due to convergence, this comparison could provide evidence that at least some of the languages of the Lesser Sundas and Moluccas subgroup with some of the languages of Sulawesi. The problem with drawing this conclusion is that sporadic lowering of high vowels (\*i > e and \*u > o) is recurrent in these languages, as shown in table 2; data are given there only for Wolio and Uma.

Wolio exhibits similar irregularities in a number of loanwords from Malay/Indonesian. While the final vowel of *kampo* ‘village’ or *manjo* ‘cup’ can be attributed to conditioned lowering in Malay (*kampon*, *manjok*), no such explanation is available in loans such as *genco* ‘lipstick’ (Malay *gincu* ‘rouge, lipstick’, ultimately from Chinese), *gola* ‘sugar’ (Malay *gula*), *konduru* ‘pumpkin’ (Malay *kundur* ‘wax gourd’), *mori* ‘pupil, student’ (Malay *murid*), or *poasa* ‘fasting’ (Malay *puasa*, from Sanskrit). Although shared sporadic changes must be given serious consideration as subgrouping evidence, then, given the wider pattern of high vowel lowering in Wolio, Uma, and no doubt other languages of central and southeast Sulawesi, this shared irregularity must be attributed to convergence. The correctness of this conclusion is driven home by Moken *ketam* ‘black’, with \*i > e (cp. \*b-in-ahi > *binay* ‘wife’, \*ikuR > *iku:n* ‘tail’, \*nipis > *nipih* ‘thin, of materials’, \*piliq > *milik* ‘choose’). If an irregular development from \*qitəm to \*qetəm is taken at face value as evidence for a subgroup of AN languages, then that subgroup would have to include at least Moken/Moklen, many of the languages of central and southeast Sulawesi, and at least some of the languages that have previously been classified as CEMP. For various reasons, this is implausible, and closer inspection shows that Moken, like many of the languages just considered, irregularly lowers high vowels in a

TABLE 2. EVIDENCE OF RECURRENT UNCONDITIONED LOWERING OF HIGH VOWELS IN WOLIO AND UMA

PMP	Wolio	Uma	PCEMP	
*ma-qitəm	maeta	mo?eta	*ma-qitəm/ma-qetəm	‘black’
*ma-ilu	maelu	ilu	—	‘orphan(ed)’
*dilaq	dela	diila?	—	‘tongue’
*qipil	epi	—	*qipil	‘ <i>Intsia bijuga</i> ’
*Ribu	rewo	—	—	‘thousand’
*sijem	sea	—	—	‘ant’
*itu	itu	etu	*itu (?)	‘that (near hearer)’
*kiday	kere	—	—	‘eyebrow’
*hinuq	—	enu?	*inu	‘necklace’
*diqdiq	—	dede?	—	‘bubble, boil’
*ma-hiaq	maea	ea?	*mayaq	‘shy, ashamed’
*susuq	kasoso	—	*susuq (?)	‘snail’
*huni	oni	—	*oni/uni (?)	‘voice, sound’
*qunap	ona	—	*qunap	‘fish scale’
*qutek	oto	—	*otek/utek	‘brain’
*qubi	owi	—	*qubi	‘yam’
*tutuqu	toto	—	—	‘true, correct’

number of forms: \*likud > *leko:t* ‘back (anat.)’, \*lipən > *lepan* ‘tooth’, \*hikan > *eka:n* ‘fish’, \*lima > *lemaʔ* ‘five’, \*um-utaq > *mota:k* ‘to vomit’, \*hulaR > *ola:n* ‘snake’, etc.<sup>11</sup> These are hardly the only Austronesian languages that show apparently unconditioned high vowel and mid vowel reflexes of PMP \*i, \*u. It is well known that several of the Malayic languages show a similar split, and in none of these is the \*i of \*qitəm lowered (Adelaar 1992:42–48). Given these observations, the distribution of reflexes of \*qitəm that show an irregularly lowered \*i is therefore best attributed to convergence in languages that show a more general pattern of sporadic high vowel lowering.

If the lowering of \*i in reflexes of \*ma-qitəm ‘black’ cannot be used to justify a linguistic subgroup that includes various languages of Sulawesi and some or all members of CEMP, we are naturally left with the question whether this change still has any value for the CEMP hypothesis. The probabilities here are difficult to assess. Whereas languages such as Wolio, Uma, or Moken show a fairly high frequency of apparently irregular high vowel lowering, the languages assigned to CEMP show relatively little sporadic lowering. This would favor a claim that the lowering of \*i in eastern Indonesian and Oceanic reflexes of \*ma-qitəm is the product of a single historical change, but the matter remains open to question. In conclusion, then, it probably is best to treat this comparison with caution, and so not include it as evidence for CEMP.

**3.4.6 PMP \*maRi > PCEMP \*mai ‘come’.** D&G claim that a number of languages outside eastern Indonesia reflect \*maRi irregularly with \*R > Ø, and that some languages in eastern Indonesia reflect \*maRi with a nonzero reflex of \*R. For the latter, they cite Buru *mahi*, Watubela *go-mari*, Geser *mari* ‘come’, but the latter two forms appear to be erroneous, as Watubela has *elat*, and Geser *ratan* ‘come’ in the Austronesian Basic Vocabulary Data Base, a source that they claim to have used (2008:114, fn. 2). Buru *mahi* does regularly reflect \*maRi, as does Arosi *mari* (next to *mai*) ‘hither, come hither’, and so, at the very least, this would be an innovation of the general form \*A > A ~ B.

However, D&G claim further that \*maRi shows an irregular loss of \*R in a number of the languages of Sulawesi, and in Yami of Botel Tobago Island. In fact, Yami *mayi* is a regular reflex of \*maRi (Tsuchida, Yamada, and Moriguchi 1987), but Gorontalo *mayi* does show sporadic \*R > Ø, a change that is also reflected in the Eastern Gorontalic languages Atinggola, Bolango, Bintauna, Suwawa, and Lolak (Usup 1981:170). The reflexes of \*R in Celebic languages, however, are less clear. D&G specifically mention Tukang Besi, Kaimbulawa, Cia-Cia, and Muna. Little comparative data are available for the first three languages, but for Muna, van den Berg (1991:15) suggests that \*maRi > *mai* is a regular conditioned change. For Popalia (a dialect of Tukang Besi), \*R normally became *h*, but has disappeared in at least two common forms besides *mai* ‘come’: PMP \*ma-Ruqanay > *moʔane* ‘male, man’, \*baqeRu > *foʔou* ‘new’ (the latter with automatic

11. Drawing on a 200-word Swadesh list collected by the anthropologist Antonio Guerreiro and made public in the Austronesian Basic Vocabulary Data Base (<http://language.psy.auckland.ac.nz/>), D&G (2008:fn. 29) also cite *maedaj* ‘black’ from Punan Kelai, a small nomadic group of interior Kalimantan, and this isolated comparison leads them to some quite bizarre speculations regarding possible contact or genetic relationship between this group and other languages showing the change \*ma-qitəm > ma-qetəm ‘black’. But there is no reason to assume that this form is cognate, since \*t and final nasals evidently remained unchanged: \*mata > *mta-n* ‘eye’, \*utaq > *tuʔ* ‘vomit’, \*batu > *wutaw* ‘stone’, \*bulan > *wulun* ‘moon’, \*itam > *tam* ‘1PL.INCL’ (cp. Uma Juman Kayan *itam* ‘1PL.INCL’).

insertion of glottal stop between identical vowels). Since Gorontalo is a Philippine language, \*R >  $\emptyset$  in this morpheme must reflect more than one historical change within Sulawesi. However, if the languages of eastern Indonesia fall into several diverse groups, as D&G suggest, it would be difficult to explain why \*maRi > *mai* is almost universal in eastern Indonesia and in Oceanic, since this would imply multiple independent changes that fortuitously targeted the same morpheme. It is possible that this change was determined at least in part by frequency: \*kami '1PL.EXCL' is irregularly reflected without the medial nasal in many daughter languages due to the high text frequency of personal pronouns, and \*maRi may show a similar sporadic contraction as a high frequency main verb that could also have been used in an auxiliary function from fairly early times. Considering everything, then, it seems most likely that PMP \*maRi became PCEMP \*mai ~ \*maRi (the latter pronunciation apparently restricted and rare), and that the sporadic loss of \*R in this form occurred independently in Gorontalic languages and perhaps in Proto-Celebic or one of its prehistoric descendants.

**3.4.7 PMP \*tudan > PCEMP \*todan 'sit'.** The comparative evidence for \*tudan is rather limited, and I will abandon my claim that it provides support for CEMP.

**3.4.8 PMP \*inum > PCEMP \*unum 'drink'.** D&G recognize (1) that the change \*inum > \*unum is not a replacement innovation, and (2) that it is very common in Oceanic languages, and somewhat less common in eastern Indonesia, but is virtually absent from the AN languages of the Philippines and western Indonesia. Because they dismiss \*A > A ~ B type comparisons such as \*uliq > *uliq* ~ *oliq* as not exemplifying an innovation, it is surprising that they did not adopt the same tactic here and simply dismiss the evidence, as many languages in both eastern Indonesia and the Oceanic subgroup reflect PMP \*inum with no change to the first vowel.

There are essentially two ways to interpret a distribution of this type. First, we could assume that all or most examples of \*inum > *unum* are products of historically independent changes. Second, we could assume that this change had already taken place in the ancestor of the languages that have been assigned to a CEMP subgroup, producing variants \*inum ~ \*unum, with one or the other surviving in any given daughter language. The first interpretation is tempting, given the possibility of convergent assimilation of the first vowel to the second. The problem with this interpretation is that reflexes of PMP \*inum are common in the Philippines and western Indonesia, and the change \*inum > *unum* is unknown throughout this region. If \*inum > *unum* is due to numerous independent changes rather than to the rise of variant forms \*inum/unum in a putative PCEMP, it is difficult to explain why a similar recurrent tendency to assimilation did not take place in the languages of the Philippines and western Indonesia. D&G note that \*inum > *unum* is quite common in SHWNG and OC languages, but is less common in the languages of the Lesser Sundas and Moluccas south of Halmahera. Cases that are known from languages that have been assigned to CMP include Kambera *unuŋu*, Mamboru (western Sumba) *umu*, Ngaibor (Aru islands) *n-un*, Paulohi *umu*, and Amahai *yumu*. Since \*inum is reflected with a front vowel in Kodi, Anakalang *inu* 'drink', it is clear that the Kambera and Mamboru changes are not products of a privately shared change. The facts are less clear for Paulohi and Amahai, so minimally we can be sure that at least four, and perhaps

five, languages in eastern Indonesia show \*i > u in reflexes of this form that are not due to a change in a restricted subgroup ancestor.

D&G clearly are concerned with the asymmetry in distribution of *unum*-type reflexes: why are these absent in the Philippines and western Indonesia, not rare in languages of eastern Indonesia, and quite common in Oceanic languages? In an attempt to find an explanation for the Oceanic cases they cite Lynch's (2002:314) reconstruction of POC \*m<sup>w</sup>inum, \*m<sup>w</sup>inum<sup>(w)</sup>-i 'drink' as support for an argument that the recurrent change of \*i to u in this form was triggered by a contiguous labiovelar nasal. However, there are problems with this explanation. First, while the change \*mu > m<sup>w</sup>i or \*um > im<sup>w</sup> is recurrent in Oceanic languages (POC \*muri > Chuukese, Puluwat m<sup>w</sup>iri-; Mokil m<sup>w</sup>eri- 'behind, after', or POC \*Rumaq > Mota, Baetora, Valpei im<sup>w</sup>a; Namakir na-im<sup>w</sup>; North Tanna n-im<sup>w</sup>; Tasiriki ima and Uripiv na-im [with subsequent merger of \*m and \*m<sup>w</sup>]; Chuukese iim<sup>w</sup> 'house', im<sup>w</sup>a-n 'his/her house'; Pohnpeian iim<sup>w</sup> 'house, building', im<sup>w</sup>e-n 'house of'), the change \*m<sup>w</sup>i > mu is not. In other words, there is considerable evidence that rounding has been transferred from \*u to adjacent labial consonants in the history of Oceanic languages, but little if any evidence that the rounding of a labiovelar consonant has been transferred to an adjacent \*i. Second, Lynch's reconstruction is based on the striking agreement of Bipi m<sup>w</sup>in, Raga m<sup>w</sup>inu with supporting evidence from Lonwolwol muen, mu-initial forms in other languages of central Vanuatu such as Namakir munum, and Standard Fijian ηunu, Wayan ηinu, all meaning 'to drink'. However, Wayan normally reflects earlier \*m<sup>w</sup> as η<sup>w</sup>, as in \*m<sup>w</sup>ane > η<sup>w</sup>ane 'sibling or close cousin of opposite sex', \*m<sup>w</sup>ata > η<sup>w</sup>ata 'snake', and it is not at all clear whether the Fijian forms reflect earlier \*m<sup>w</sup>- or \*η-. Neither is it clear that the reconstruction of POC \*m<sup>w</sup>inum, \*m<sup>w</sup>inum<sup>(w)</sup>-i is justified, as the Bipi and Raga forms could have developed from \*munu, or even \*minu (cf. \*Rumaq > Raga im<sup>w</sup>a 'house', \*ma-taqu > m<sup>w</sup>atua- 'right side'). I have field data for 26 languages of the Admiralty islands, and Bipi is unique in reflecting \*inum with an initial labiovelar nasal. Nearly all other languages within the Admiralty subgroup reflect POC \*inum as in, imu, or un. Third, there is virtually no other evidence that PAN \*-um- 'actor focus' survived, as an active affix, in POC. Reflexes of POC \*inum occur with no initial consonant in the great majority of OC languages. Within Vanuatu, apparent reflexes of \*inum that contain m- are fairly common. However, it is by no means clear that this consonant reflects PAN \*-um-. Finally, even if this rather questionable scenario is accepted to explain the frequent occurrence of \*i > u in Oceanic reflexes of POC \*inum, it does not help to explain the similar recurrent pattern of irregular change in the languages of eastern Indonesia. All in all, then, it appears most reasonable to assume that PMP \*inum became \*inum ~ \*unum in a language that was immediately ancestral to most or all of the languages of eastern Indonesia and to Oceanic, namely CEMP.

To the foregoing examples of sporadic sound change that are mentioned in Blust (1993) we can add the irregular lowering of \*i in reflexes of PMP \*inep 'lie down to sleep'. The core evidence is as follows: (CMP) Selaru enaf, Koiwai ena-n 'sleep, lie down', Watubela g-ena 'sleep', (SHWNG) Weda yenef, Numfor enaf 'sleep', Windesi ena 'sleep' ena-mia 'dream', Serui-Laut ena, Waropen ena-ko 'sleep', (OC) Wogeo o-yeno, Manam eno 'sleep, lie down', Gedaged en 'lie, rest, sleep', Dobuan ?eno 'sleep', ?eno?eno 'lie down', Kilivila kenu- 'lie down; go to sleep', Misima keno-keno 'sleep', keno-nou

‘dream’, Motu *eno*, Nauna *en*, Mengen *keno*, Sa’a *eno(eno)*, *enoh-i*, Raga *eno*, Nguna *one* (met.) ‘lie down’. Again, this is an innovation of the form \*A > A ~ B, since some putative descendants of PCEMP do not lower \*i, as with Buru *ine*, Raluana *inep* ‘lie down, sleep’, Kairiru *yin* ‘recline, lie down, settle (residentially)’, and again it is a lexically specific sound change, because \*i normally does not lower to a mid vowel in any of these languages.

**3.4.9 Prefixal/proclitic agreement on the verb.** The history of prefixal or proclitic agreement marking on the verb is one of the least understood aspects of AN comparative linguistics. Wolff (1996) has argued that this type of structure arose from a generalization of structures in which an enclitic pronoun such as *ku* ‘1SG’ was preposed to the verb when the predicate was initiated by a modifier or auxiliary. D&G (2008:132) note that such structures are widespread, including many languages in eastern Indonesia, a number of the languages of western Melanesia, and a scattering of languages in Sulawesi, Sumatra, and the Mariana islands (Chamorro). If Wolff is correct (and the evidence suggests that he is), then the use of prefixal/proclitic agreement markers on the verb has arisen independently through multiple historical changes. This obviously weakens it as subgrouping evidence, although it is still noteworthy that systems of this type are extremely common in eastern Indonesia, and are found in a number of the languages of western Melanesia.

**3.4.10 Alienable/inalienable possessive distinction.** D&G (2008:142) argue that, although distinctions between alienable and inalienable possession are found in eastern Indonesia, “it is only in EMP that they become the norm, and where cognate morphology can regularly be found.” This point is well taken, since without cognate morphology the comparison relies on typological data for subgrouping, and it is well known that this is unreliable. Nonetheless, the reported occurrence of a distinction between edible and general possession for alienable relationships in Selaru is striking, since such a semantic distinction in possessive marking is unknown among AN languages outside eastern Indonesia and the Oceanic group.

### 3.4.11 Frozen morphology

**a. PMP \*hepat > PCEMP \*həpat, \*pat, \*pati ‘four’.** D&G (2008:143–44) challenge these proposed innovations as evidence for CEMP on the grounds that a number of the languages of Sulawesi have also lost the initial vowel of PMP \*hepat. But this ignores the wording of the original text: “Although many CMP languages and all Oceanic languages show an unexplained loss of the first-syllable vowel of PMP \*hepat, a few CMP languages retain this vowel. More distinctive is the appearance of a final \*i in a number of CMP languages reported in Wallace (1962), and in many Oceanic languages, particularly in the Solomons and Vanuatu.” Examples include (CMP) Wahai *ati*, Teluti *fai* ‘four’, (SHWNG) Marau *ati*, Munggui *bo-ati*, Dusner *ati ~ pati* ‘four’, and (OC) Manam *wati*, Kilivila *-vasi*, Mono-Alu *e-hati*, Nggela *vati*, Arosi *hai*, Raga *fasi*, and Nguna *pati* ‘four’. Blust (1995) contains a six-page entry documenting reflexes of PAN \*Sepat ‘four’, including nineteen affixed forms of this stem that can be reconstructed for protolanguages ranging from PAN to Proto-Philippines, and reflexes of \*pat-i are confined to CMP, SHWNG, and Oceanic languages.

**b, c. PMP \*ma-huab > PCEMP \*mawab ‘yawn’, PMP \*ma-hiaq > PCEMP \*mayaq ‘shy, ashamed’.** These two comparisons are treated separately by Blust (1993)

and by D&G, but can be combined, as they exemplify essentially a single change. D&G (2008:144) argue that since Palu'e *yoa* 'to yawn', and Buru *duba* 'sleepy, yawn' appear to reflect \*huab, the change \*ma-huab > \*mawab in not universal in the collection of languages that have been called "Central-Eastern Malayo-Polynesian." There are a number of problems with this objection.

To begin with *duba*, the reader is referred to an unpublished paper by Grimes for arguments that this rather improbable etymology is in fact valid. Devin 1989, which remains the best source of lexical information available for Buru, gives *duba*, *dub-duba*, *duba-duba* as 'sleepy, drowsy', and *dubat* as 'drowsiness, sleepiness', but does not mention 'to yawn' as an associated meaning, citing instead *mawa* (vi) 'to yawn'. Even if we were to accept Buru *duba* as meaning both 'sleepy' (stative) and 'to yawn' (active), there are serious problems in trying to derive this form from PMP \*huab. First, D&G speculate that the initial consonant of *duba* may be a fossilized agreement prefix \*da- '3PL', but cite no parallel cases, and give no explanation as to why a plural prefix would be fossilized on a verb stem that typically involves a singular actor. Second, in an effort to reduce the implausibility of the comparison, they assume a sporadic metathesis between the final two segments. This in itself weakens the claim for cognation, since it essentially introduces an irregularity into the sound correspondences, and so increases the role of chance in producing a match between etymon and hypothesized reflex. As it happens, this particular hypothesis is even more problematic, since -VC metathesis is vanishingly rare in AN languages generally (Blust forthcoming, table 9.11). D&G are not explicit about the Buru reflexes of PMP \*b, but comparisons such as PMP \*babuy > *fafu* 'pig, boar', \*baŋuŋ > *faju* 'arise, wake up', \*betaw > *fetu* 'sister of a man', \*biRaq > *fiha* 'wild taro', \*bulan > *fulan* 'moon, month', \*tuba > *tufa* 'derris root fish poison', \*bibiR 'lip' > *fifi-n* 'lips, mouth', or \*bulawan > *eflawan* 'gold' show clearly that PMP \*b normally became Buru *f* in nonfinal position. As noted by Collins (1982:84), the change \*b > *f* is found in all known members of the Sula-Buru subgroup, and *b* in these languages reflects PMP \*mp/mb. It is most likely, then, that *duba* reflects an earlier form \*dumpa or \*dumba. It might still be possible to salvage the etymology \*huab > *duba* if metathesis followed the change of nonfinal \*b > *f*. But this in itself depends upon the reflex of \*b in word-final position. As is well known, reflexes of word-final \*b are often difficult to find in AN languages, but since Buru has *mawa* 'to yawn' (< PMP \*ma-huab), this suggests that \*-b was lost. We are left, then, with an even narrower range of possibilities: since \*b > *f* must be restricted to nonfinal position, metathesis of \*d-uab > *duba* had to occur after this change, and loss of final \*b had to occur last. On balance, the etymology that D&G propose presents more problems than advantages in understanding the issue at hand, and the wisest course of action would be to abandon it.

Despite its initial appeal, the citation of Palu'e *yoa* 'yawn' as a reflex of \*huab is nearly as problematic as Buru *duba*. Unpublished, but rather extensive, vocabularies of a number of the languages of the Lesser Sundas collected by the anthropologist Margrethe Dirkwager in the early 1980s include data for Palu'e, Bha'i (Béna), Haga (Méli), Lio, and Kédang. The word for 'to yawn' in these five lists is: Palu'e *yoa*, Bha'i (Béna) *moapa*, Haga (Méli) *moapa*, Lio *moa*, and Kédang *moa*. It is quite possible that all of these forms except Palu'e *yoa* derive from \*mawap via an intermediate step \*moap (cf.

PMP \*ma-quhaw > *mauw* > *mawaw* > Haga [Méli], Bha'i [Béna] *moa* 'thirsty', PMP \*taqun > Palu'e *to* 'year'). Although Palu'e *yoa* cannot be reconciled with \*mawap, it may very well have been a reformation based on \*mawap, since \*u > o is irregular. Regardless of how this matter is settled, the vast majority of CEMP reflexes point to \*mawap, and D&G are therefore insisting once again—contrary to the working practice of most historical linguists—that changes of the general form \*A > A ~ B are not innovations. The only real issue here with regard to subgrouping is how to assess the probability that \*ma-hiaq > \*mayaq was a recurrent change involving multiple historical events, or a single change in a language ancestral to most or all of the languages of eastern Indonesia and to Oceanic.

Finally, D&G remark, somewhat gratuitously, that “the change from \*ma-huab > \*mauab > \*mawab is primarily one of transcription, and not of phonemicization (see Clynes 1997, 1999 on the status of ‘diphthongs’ in Austronesian reconstructions).” But if \*u and \*w were different phonemes in PMP, as is normally assumed, then a change from \*mauab (a trisyllable) to \*mawab (a disyllable) is clearly a phonemic change, not simply a matter of “transcription.”

As for PMP \*ma-hiaq > PCEMP \*mayaq 'shy, ashamed', D&G (2008:145) suggest that Soboyo *mahi* 'shy, ashamed' reflects \*ma-hiaq without loss of \*h and resyllabification. However, Soboyo does not normally lose final vowels. Moreover, Blust (1981:25) considered this comparison within the wider context of Soboyo historical phonology, and concluded that *mahi* was best treated as a borrowing of Ternate *mahe*, a fact that D&G pass over in silence.

**3.4.12 PMP \*t-ina 'mother' > PCEMP \*t-ina 'big'.** With regard to this proposed innovation, D&G (2008:145) object that the sense of 'larger member of a set' is associated with a reflex of \*t-ina in Malay and a number of the languages of Sulawesi, and they cite examples to make their case ('bow' in relation to 'arrow', 'mortar' in relation to 'pestle', 'thumb' in relation to other fingers, etc.). They are correct in this observation, but they miss the point of the original claim, which is that a reflex of PMP \*t-ina does not just refer to the larger member of a set, but actually means 'big' in a number of the languages that have been classified as CEMP. This is particularly clear in Jonker (1908), who gives separate dictionary entries for Roti *ina* 'wijffe, vrouwelijk van dieren, wanneer zij volwassen zijn' (female of animals, when they are full-grown), and *ina* 'groot ... in samenstelling met allerlei substantieven' (large ... in combination with all sorts of nouns). As examples of the latter usage he cites *uma-ina* 'a big house', *soka-ina* 'a big sack', *musa-ina* 'a big (is)land', *uda-ina* 'heavy rain', and *leledo-ina* 'the whole day', clearly a far more radical extension of the original sense of this term than those that D&G cite from Malay or various languages of Sulawesi, all of which involve larger or more important members of a set in relation to smaller or less important members. Mathijssen (1906) reports a similar usage for Tetun *inan*, which he treats lexicographically as a single polysemous form rather than as two homophonous entries: 'female of animals; many'. Examples of usage include *emma inan* 'many people', *inan basuk* 'very many', *nia mai loro sei inan* (3SG come sun still intensive) 'He came while the sun was still high in the sky/while it was still bright daylight', *tihar inan* 'a big drum'. In addition he gives *inan aman* (< \*ina 'mother', \*ama father') 'much, many (usually only *inan*)' (cp. *ina ama* 'parents'). Reflexes of PMP \*t-ina

‘mother’ are used in a very similar way in some Oceanic languages, as Bipi, Lindrou *tinan*, Sori *tinay* ‘big’ (cp. Bipi, Lindrou *tine-n*, Sori *tiney* ‘his/her mother’, the two reflexes of PMP \*t-ina now lexically dissociated in all three languages). For Nggela of the central Solomons, Fox (1955) gives *tina* with three meanings: 1. mother, 2. a woman of her standing in one’s clan, 3. large of its kind. Examples of the latter sense that he cites include *mbeti* ‘water, stream’, *mbeti tina* ‘river’, and *pari* ‘soil, earth, ground’, *pari tina* ‘mainland, land’. In addition, he cites *tina* as marking the meaning ‘mature’, as in *na gai te tina tua* (ART tree at/by mature already) ‘the tree is grown already’, and ‘past, last’, as in *miulu tina* ‘last year’. For Manam, Böhm (1975) gives *tina* ‘mother, aunt’, and then separately *tina* (adv.) ‘very’. While expressions such as ‘mother of waters’ for ‘river’ may be products of language universal tendencies in the use of metaphor, the adjectival and adverbial uses of a reflex of PMP \*t-ina ‘mother’ in various languages of eastern Indonesia and in Oceanic appear to be diagnostic of particular historical developments that are not shared with other AN languages. The suggestion that these comparisons be dismissed on the grounds that similar semantic extensions are likely to arise independently, or have arisen independently in WMP languages, has therefore not been convincingly demonstrated.

**3.4.13 PMP \*m-udehi > PCEMP \*mudi ‘back (anat.)’.** PMP \*m-udehi covered a wide semantic range, including at least ‘last; come after or behind; late, later; future; stern of a boat; youngest child’ (Blust 1995), but no known reflex of this term means ‘back (anat.)’ in a non-CEMP language. Reflexes of PAN \*likud ‘back (anat.)’, on the other hand, are fairly common in this meaning in Taiwan, the Philippines, Borneo, and the Philippine languages of Sulawesi, but are rare elsewhere in western Indonesia, and disappear entirely in eastern Indonesia and the Pacific: (Taiwan) Thao *rikus*, Tamalakau Puyuma *rikuz-an*, Saaroa *likusu*, (Philippines) Ilokano *likúid*, Tagalog *likód*, Binukid *ta-likud-an*, Tausug *ta-ykud*, (Borneo) Kadazan *hikud*, Bintulu *likud*, Long Anap Kenyah *likut*, Katingan *likur*, (Burma-Thailand) Moken *leko:t*, (Sulawesi) Sangir *likudə?*, Tontemboan *likud ~ likur*, Bare’e (Pamona) *ta-liku*, Bolaang Mongondow *likud*, Wolio *ta-liku* ‘back (anat.); behind’. Bolaang Mongondow *likud* and Wolio *ta-liku* ‘back (anat.); behind’, are described as having both anatomical and nonanatomical references, but all other terms are described in the primary sources as nouns that refer to the back of humans or animals. Blust (1993:261) noted that “PMP \*m-udehi evidently persisted as PCEMP \*mudi ‘behind, afterwards’, but in the meaning ‘back (anat.)’ it apparently replaced PMP \*likud, which survives only in occasional reflexes of the morphologically complex form \*ta-likud ‘turn the back on.’”

D&G (2008:146) challenge this claim on several grounds. First, it is asserted that “Palu’e *kuri* ‘behind’ reflects \*likuD via a metathesized form \*kuliD (or, with more convolutions, \*kuDil). This clearly indicates that PMP \*likuD was not replaced by \*mudi, as Blust (1993:261) suggests, and that (at least in the west of the CMP area) \*likuD was retained.” Second, D&G claim that the exclusively shared semantic change seen in PMP \*m-udehi ‘afterwards, behind’ > PCEMP \*mudi ‘back (anat.)’ is invalid as subgrouping evidence because (1) some reflexes of PMP \*m-udehi in languages of eastern Indonesia still mean ‘afterwards, behind’, and (2) some apparent reflexes of \*likud in languages of eastern Indonesia have a similar meaning, as Dhao *li?u* ‘1. behind (location), 2. outside’, or Helong *liku-n* ‘behind, outside (e.g., a house)’.

With regard to the first objection, the claim that Palu'e *kuri* 'behind' reflects \*likud assumes syllable metathesis, a phonological process for which little, if any, unambiguous evidence exists in the literature on phonological change in natural languages. Moreover, there is no precedent for the change \*l > r in Palu'e, which D&G assume on a purely ad hoc basis (cf. \*laki > laki 'male', \*lima > lima 'five', \*qulej > hulé 'caterpillar', \*qilih > ili 'mountain', \*telu > təlu 'three', \*bulu > wulu 'body hair', \*salaq > sala 'error', etc.). Given both its meaning and its form, Palu'e *kuri* is far more likely to reflect a prefixed form of \*udehi similar to Cebuano *ka-ulahi-an* 'the last', or Minangkabau *k-udi-an* 'afterwards' (Blust 1995), since \*d, \*j, and \*z generally merged as -r-: \*ma-quḍip > moré 'living, alive', \*suduk > huru 'spoon, ladle', \*tindes > tiro 'crush lice', \*ḥajan > ḥara 'name', \*qapeju > pəru 'gall (bladder)', \*quzan > hura 'rain'.

With regard to the second objection, it is apparent once again that D&G are confused about the nature of linguistic innovations. As noted above, Blust (1993:261) clearly recognized that PCEMP \*mudi continued to mean 'afterward, behind', but in addition had acquired the sense 'back (anat.)', which had not previously been part of the denotation of this term. This is thus another example of an innovation that takes the general form \*A > A ~ B. The further citation of forms such as Dhao *liḡu* '1. behind (location), 2. outside', or Helong *liku-n* 'behind, outside (e.g., a house)' has no clear relevance to the argument, since the claim is that a reflex of PMP \*ma-udehi replaced \*likud in the meaning 'back'. The only real issue is therefore whether this semantic innovation is more likely to have happened once in a single language ancestral to the hypothesized CEMP subgroup, or is a product of multiple historical changes that produced convergent results. Since reflexes of PMP \*udehi (and its affixed variants) are fairly common in the languages of the Philippines and western Indonesia, the absence of a parallel semantic change in these languages clearly favors the hypothesis that the change \*m-udehi 'afterward, behind' > \*mudi 'back (anat.)' in languages of eastern Indonesia and the Pacific is the product of a single innovation.

Finally, D&G note that "Southern Tukang Besi *mburi* 'behind' shows the same fusion of \*m- with the root seen in the CEMP-area languages, and the same loss of the \*-eh-. At the same time, the northern Tukang Besi form *taliku* 'behind, back' obviously reflects \*ta-likuD, and not \*likuD alone, indicating that this innovative form is not confined to CEMP-area languages, and suggesting a fused preposition \*ta, not otherwise present in the language." Reflexes of PMP \*ma-udehi (written \*m-udehi in Blust 1993:261) are in fact widespread in languages of the Philippines and western Indonesia outside Sulawesi: Yami *ma-udi*, Itbayaten *ma-wdi* 'stern of a boat', Isneg *m-udi* 'last, hindmost', Ifugao *ma-udi* 'it will be made last', Bikol *ma-huri* 'be last, be late', Kelabit *m-ud<sup>h</sup>ih* 'behind, last', Malay *udi* 'posterior; hindward; subsequent', *kem-udi* 'rudder of a boat', Sundanese *juru mudi* 'helmsman, steersman' (Blust 1995). Given this readily available observation, the relevance of Tukang Besi *mburi* to the claim at issue is obscure. Likewise, reflexes of PMP \*ta-likud 'turn the back on' (where \*ta- is a prefix, not a preposition) are found from at least the southern Philippines (Taosug *ta-ykud* 'back') and Borneo (Kadazan *to-hikud* 'turn backwards, turn one's back to', Kayan *tə-likum* 'sit back-to-back; turn the back'), through eastern Indonesia (Yamdena *na-t-liur* 'turn the back'), to Melanesia (Tolai *taliur*

'backwards, with head facing backwards'). Why this should be termed an "innovative form" shared by *Tukang Besi* with "CEMP-area languages" is not explained.

**3.4.14 PMP \*ma-qitem 'black' > PCEMP \*ma-qetəm 'dirty'.** The sporadic vowel lowering in PMP \*ma-qitem PMP \*ma-qetəm 'black' has already been considered, but reflexes of this form also show an apparent semantic innovation in languages of eastern Indonesia and the Pacific that is not known to be shared with any other AN language.

Reconstruction of a form meaning 'dirty' for early AN protolanguages such as PAN or PMP has proved difficult. Blust (1993) suggested PMP \*cemeD, but with a question mark to indicate its uncertain status. PAN \*daki 'dirt on the skin' is a noun reflected in a number of languages from Taiwan to eastern Indonesia, but this word apparently did not occur, even in affixed form, in the general sense of 'dirty'. A number of languages from the western Lesser Sundas to the central Moluccas show an innovation in which the reflex of PMP \*ma-qitəm 'black' has come to mean both 'black' and 'dirty'. Where textual examples are available, this semantic extension appears to apply not just to the meaning 'dirty' in a sense coterminous with 'black', but rather to usages in which 'dirty' does not imply any color referent (e.g., a dirty plate): (Lesser Sundas) Ngadha *mité* 'black, dark; dirty' (Arndt 1961:332), Sika *mita-k* 'dirty, black', *gebu mitak rakaŋ* (cloud black very) 'very black clouds', *pigaŋ mita-k* 'a dirty plate' (Meyer 1964:189), Kédang *miteŋ* 'dark, black, dirty' (Samely 1991:183), Buru *mite-t* 'black; dirty' (Tryon 1995, 4:227, 264); (Vanuatu) Peterara, Wailengi *maeto*, Nasawa *maito*, Ngwatua *meto*, Labo, Windua *mite*, Bongabonga, Tongariki, Namakir, Mataso *ma?et* 'black; dirty'. A number of languages in various parts of Melanesia also reflect other bases with both meanings: (Admiralties) Ahus *rokodan*, Bipi *ŋakitan*, Pak *akhan*, (Solomons) Vaghua *poŋo*, Katazi *pito*, Nggeri *mbau*, (Vanuatu) Merlav *mbolo*, North Malo *burixa*, Sesake *loa* 'black, dirty'.

This comparison does not provide strong evidence for subgrouping, as the probability of convergent semantic innovation is fairly high. However, if it is not the product of a single historical change in which PMP \*ma-qitem 'black' became PCEMP \*ma-qetəm 'black; dirty', it is puzzling why a similar semantic extension has not yet been found in any AN language outside the hypothesized CEMP group.

**3.4.15 PMP \*zuRi 'fish bone' > PCEMP \*zuRi 'bone'.** One of the most striking lexical differences between CEMP languages and those "higher up" the AN family tree is seen in the word for 'bone'. Although other morphemes represent this meaning in individual languages, reflexes of PAN \*CuqelaN/Cuqelan 'bone' are widely distributed throughout Taiwan, the Philippines, and western Indonesia, but are unknown in eastern Indonesia and in Oceanic. Blust (1978b:111–13) noted that next to Dempwolff's \*duRi 'thorn', it apparently is necessary to posit \*zuRi 'fish bone'. Given the close formal and semantic similarity of these forms, it is not surprising that they appear to have been confused in the history of some languages. What is noteworthy is that reflexes of PMP \*zuRi occur with the general meaning 'bone' in many of the languages of eastern Indonesia and in POC \*suRi. D&G (2008:147) erroneously attribute recognition of this comparison to Mahdi (1994), and dispute the claim that \*zuRi 'fish bone' > \*zuRi 'bone' can be taken as an innovation defining a CEMP subgroup. Rather than interpreting this as a relatively straightforward example of widening of reference, they suggest instead that "as attested

in the northern Philippines and northern Sulawesi (both in the WMP area), ‘thorn’ came to take on ‘fish bone’ as an additional function; this sense continued and, possibly due to founder effects, was exaggerated in the CEMP area, with the additional erratic, but logical, extension to ‘bone’.”

This is a remarkably elaborate hypothesis to account for a lexical distribution than can be explained in a far more straightforward fashion as a semantic extension from ‘fish bone’ to ‘bone’. Moreover, it fails to explain the initial consonant of POC \*suRi ‘bone’, which was clearly distinguished from \*ruRi ‘thorn’ (Blust 1978b:111). Words for ‘bone’ in many AN languages can be applied both to fish bones and to the bones of higher-level vertebrates, but there is clearly a conceptual distinction between condyloid bones (those that fit into sockets), and fish bones, which are much more thorn-like. Wilkinson (1959:1244) lists Malay *tulaj belut* ‘eel bones’ as an expression for a herring-bone pattern, and *ikan duri* for certain sharp-finned catfish, but does not give *duri* in the meaning ‘bone’. However, more recent dictionaries of Indonesian, such as Moeliono (1989), list one definition of *tulaj* as ‘*duri ikan*’ (‘fish bone’), where the word for ‘thorn’ is used to represent a fish bone as distinct from the bones of humans, other mammals, reptiles, amphibians, or birds. In many of the languages of eastern Indonesia, \*d and \*z have merged, so that words for ‘bone’ could reflect either \*duRi or \*zuRi. D&G (2008:147) maintain that “even if we accept the Proto-Oceanic innovation \*suRi, another issue is whether the CMP-area language forms reflect the innovation \*zuRi, or whether they are semantic extensions (not innovations) on the earlier PMP \*DuRi ‘thorn’.”

In other words, D&G are willing to accept that POC \*suRi ‘bone’ is a reflex of PMP \*zuRi ‘fish bone’ with semantic extension to all types of bones, but they wish to treat all similar words for ‘bone’ in eastern Indonesia as deriving from \*duRi ‘thorn’. This interpretation is unconvincing. A semantic change from ‘thorn’ to ‘bone’ is unlikely unless it was mediated by ‘fish bone’. Moreover, Oceanic forms such as Motu *duri* ‘sting of a stingray, barb on an arrow’ show clear evidence of cross-over between reflexes of PMP \*duRi and \*zuRi, as PMP \*d became POC \*r, Motu *r*; and PMP \*z became POC \*s, Motu *d* (Ross 1988:205–6).

**3.4.16 PMP \*buhek > PCEMP \*daun ni qulu ‘hair of the head’.** Both Blust (1993:262) and D&G (2008:147–48) suggest that this innovative expression was calqued from a Papuan source. Given the presence of similar expressions in some Papuan languages, but in no AN language outside eastern Indonesia and Oceanic, this does appear to be the best available explanation. However, D&G prejudge the matter in holding that “the evidence is clearly in favor of \*daun ni qulu being a calque from Papuan languages, and not an innovation of PCEMP.” The gratuitous implication of this statement is that these alternatives are mutually exclusive: either \*daun ni qulu was calqued from a Papuan source, or it was an innovation of PCEMP. But if this is a likely calque, the point at issue is whether it reflects a single historical change in a language community ancestral to the languages of eastern Indonesia and Oceanic, or is a product of multiple contact-induced changes that produced similar outcomes. Blust (1993) opts for the former interpretation, and D&G for the latter, but there appears to be no obvious way to settle this issue on present evidence.

**3.4.17 PMP \*dalem ‘inside, interior’ > PCEMP \*daləm ‘mind, feelings’.** Blust (1993:262) noted that reflexes of PMP \*dalem ‘inside, interior’ appear in some CEMP languages with the meaning ‘mind, feelings’. It was also pointed out that indirect uses of a reflex of \*dalem with similar semantics occur in WMP languages, as with Ilokano *dalem* ‘liver’ (the traditional seat of the emotions), or Malay *dalam hati* (lit. ‘in the liver’) ‘secretly or silently’. D&G argue that direct uses of a reflex of \*dalem with similar semantics also occur in WMP languages, as with Muna *lalo* ‘heart; seat of emotions’. (Note that many idioms relating to feelings are formed with this word.) They are certainly correct in this observation, but what remains noteworthy is that a reflex of \*dalem in CEMP languages is recurrently associated not just with the emotions, but more specifically with the mind: Tetun *laran metan* (‘inside’ + ‘black’) ‘the large intestine; the inner soul, inner nature, *the intellect*’, Buru *lalen* ‘inside, heart, *mind*; seat of one’s emotions’, (OC) Motu *lalo* ‘interior, *the mind*’; *lalo-a* ‘think, remember’, Arosi *raro* ‘the inside, inner part; the feelings, heart, *mind*, disposition’, Kiribati *nano* ‘soul, conscience, heart, will, desire, sentiment’, *nano-a* ‘think about, keep in mind, be preoccupied with’ (my italics). Reflexes of PMP \*dalem meaning ‘mind’ have not been reported in any WMP language to my knowledge, and it probably would be a mistake to equate this sense with the emotions.

D&G end their criticism of this comparison with an expatiation on how reflexes of PMP \*qatay are used in many AN languages to refer to emotional states or character traits. The relevance of this digression to the point at hand is obscure, and in any case they seem unaware of the material in Blust (1995), where comparative evidence is presented for eleven affixed forms of PMP \*qatay, including ‘big liver’ = ‘brave, courageous; proud, arrogant’, ‘burning liver’ = ‘angry, furious’, ‘rotten liver’ = ‘filled with malice’, ‘sick/hurt liver’ = ‘resentful, offended’, ‘small liver’ = ‘afraid, timid’, and ‘white liver’ = ‘pure-hearted’.

**3.5 SUMMARY OF THE EVIDENCE FOR CEMP.** As noted in table 1, D&G claim that only one of the innovations proposed for CEMP in Blust (1993) stands up to closer scrutiny. Ironically, that proposed innovation is PMP \*tudan > PCEMP \*todan ‘to sit’, the one comparison that I now feel least able to justify. With regard to the remaining evidence, however, we obviously have very different views not only about particular interpretations, but about how the Comparative Method of linguistics works with regard to subgrouping and to the identification of cognates. Whereas I recognize all demonstrable innovations as carrying subgrouping weight (obviously, of varying degrees, depending upon the particulars of each case), D&G allow only replacement innovations as valid markers of subgroups, dismissing all other innovations out of hand. Moreover, where I follow the Comparative Method in insisting on recurrent sound correspondences to establish cognation, D&G repeatedly resort to phonetic similarity, supplemented by a variety of ad hoc hypotheses regarding morphology, metathesis, and unique phoneme matches. The results of research agendas guided by such different commitments to method are predictably divergent, as seen in table 3 (in the rightmost column 3 = strong support, 2 = moderate support, 1 = weak support, 0 = no support).

To summarize, where D&G find little or no support for CEMP, I find considerable support for this group. As noted above, these disagreements can be traced in large part to

differences in the working methods used: D&G accept only replacement innovations as subgrouping evidence, whereas I accept any demonstrable innovation; D&G repeatedly appeal to phonetic similarity to claim cognation, whereas I use only evidence of recurrent sound correspondences; D&G cope with nonconforming lexical evidence such as the reflexes of PCEMP \*kandoRa ‘cuscus’ and \*mans(aə)r ‘bandicoot’ by inventing a source language from which these terms could have been borrowed repeatedly by genetically disparate AN languages entering eastern Indonesia at different times and places, whereas I reject the use of such ad hoc hypotheses.

The further claims by D&G that “many of the traits ascribed to the CMP- or CEMP-area languages can be found in more conservative WMP-area or Formosan languages as well” (2008:114), that “any PCEMP node should, logically, include some languages that have been assumed to be classified as ‘WMP’” (2008:115), that “at least some of the CMP-area languages (from around Flores) share features with a subgroup of languages that have been treated as WMP, those of (Southeast) Sulawesi” (2008:116), and that “a large number of the features that have been described as characterizing CEMP languages are also found in the WMP languages that border them” (2008:117), have been found either to be baseless, or to have no subgrouping value when they are examined in greater

**TABLE 3. THE CEMP INNOVATIONS RECONSIDERED AGAIN**

<b>Lexical innovations</b>	<b>D&amp;G (2008)</b>	<b>Blust (2009)</b>
1. Marsupial terms	(dismissed)	3
2. Hawksbill turtle	(dismissed)	3
3. Other lexical innovations	(dismissed)	1–3 (varying)
<b>Regular phonological innovations</b>		
4. Cluster reduction, 1: $C_1V_1C_2 C_1V_1C_2 > C_1V_1C_1V_1C_2$	(no support)	3
5. Cluster reduction, 2: $CVN_1C_1VC > CVN_1C_1VC$	(no support)	(irrelevant)
<b>Irregular phonological innovations</b>		
6. PMP *uliq > PCEMP *oliq ‘return’	(no support)	2
7. PMP *i-sai > PCEMP *i-sei ‘who’	(no support)	1
8. PMP *ma-qitəm > PCEMP *ma-qetəm ‘black’	(no support)	?
9. PMP *maRi > PCEMP *mai ‘come’	(no support)	2
10. PMP *tudan > PCEMP *todan ‘sit’	(yes)	0
11. PMP *inum > PCEMP *unum ‘drink’	(no support)	2
12. PMP *inēp > PCEMP *enəp ‘sleep’	—	2–3
<b>Morphosyntactic innovations</b>		
13. Prefixal/proclitic agreement on verb	(no support)	1
14. Alienable/inalienable possession distinction	(no support)	1–2
15. Frozen morphology		
a. PMP *həpat > PCEMP *həpat, *pat, *pati ‘four’	(no support)	3
b. PMP *ma-huab > PCEMP *mawab ‘yawn’	(no support)	2
c. PMP *ma-hiaq > PCEMP *mayaq ‘shy’	(no support)	2
<b>Semantic innovations</b>		
16. PMP *t-ina ‘mother’ > PCEMP *t-ina ‘big’	(no support)	2–3
17. PMP *m-udehi ‘behind’ > PCEMP *mudi ‘back (of body)’	(no support)	2–3
18. PMP *ma-qitəm ‘black’ > PCEMP *ma-qetəm ‘dirty’	(no support)	2
19. PMP *tuqəla(nj) > PCEMP *zuRi ‘bone’	(no support)	3
20. PMP *buhək > PCEMP *daun ni qulu ‘head hair’	(no support)	?
21. PMP *daləm ‘inside’ > PCEMP *daləm ‘mind, feelings’	(no support)	2

detail.<sup>12</sup> A number of similar claims were made by Nothofer (1992), and answered in Blust (1993), but the idea continues to reappear that there is a subgrouping connection between some or all of the languages of southeast Sulawesi and various of the languages of eastern Indonesia. The fact remains that no convincing evidence has ever been presented to justify such a claim.

**4. CENTRAL MALAYO-POLYNESIAN REVISITED.** Blust (1993:263) took pains to stress that CMP and CEMP are not equivalent types of genetic unit: whereas CEMP is an “innovation-defined” group in the terms of Ross (1988, 1997), CMP appears to be an “innovation-linked” group. What Ross has tried to capture in this terminological distinction is a difference between two generically different configurations of historical change in language split. In the first of these, language communities separate sharply by means of geographical or social distance into descendants that (1) continue to change in relative isolation from one another, and (2) remain geographically compact, at least during the initial period of separation. As a result, innovations accumulate over time that are shared only by languages within each branch, and such groups are clearly defined by collections of exclusively shared innovations. In the second configuration, a language community divides and almost immediately spreads over a considerable geographical range to form a dialect chain. Innovations that originate in any part of such a chain then spread easily through portions of the chain, but rarely extend throughout its entire length. The result is a collection of languages with many innovative features shared by some, but not all, members of the putative group. CMP appears to be a group of the latter type, an innovation-linked subgroup, or “linkage” as defined by Ross.<sup>13</sup>

Needless to say, this terminological distinction provides only a rough set of guidelines for distinguishing genetic groupings of languages that may also differ in other ways. As noted in Blust (1978a), for example, the South Halmahera-West New Guinea group evidently began as a dialect chain, since the loss of final vowels was innovated in southern Halmahera and spread eastward, while loss of \*k was innovated in Cenderawasih Bay and spread westward, the two changes overlapping in different orders in languages such

12. Although these statements evidently are intended as claims about linguistic subgrouping, they are actually quite vague. Subgroups are based on evidence of exclusively shared innovations that are not likely to be products of convergence, and statements such as “at least some of the CMP-area languages (from around Flores) share features with a subgroup of languages that have been treated as WMP, those of (Southeast) Sulawesi” say nothing about how these shared features were acquired. In fact, at one point D&G (2008:117) propose an entirely different explanation: “The fact that a large number of the features that have been described as characterizing CEMP languages are also found in the WMP languages that border them is strongly suggestive of diffusion, because the features are not confined to a particular subgroup.” Needless to say, this interpretation is at odds with the tree they propose, which is reproduced here as figure 2.

13. D&G (2008:115) have a different conception of language split, since they speak of “a linkage, rather than an undifferentiated protolanguage.” This statement implicitly adopts the questionable assumption that some protolanguages are dialect-free, and at the same time implies that a linkage cannot descend from “an undifferentiated protolanguage.” But in principle there is no obvious reason why a fairly homogeneous protolanguage could not expand rapidly and thus give rise to a linkage. In other words, the difference between a discrete subgroup and a linkage does not reflect different types of protolanguages so much as different relationships between the passage of time and the movement into new territories (little movement over long periods of time for discrete subgroups, much movement over short periods of time for linkages).

as Buli of south Halmahera (with  $*-V > \emptyset$ , then  $*k > \emptyset$ ) and Numfor of Cenderawasih Bay (with  $*k > \emptyset$ , then  $*-V > \emptyset$ ). Today the communities that share these overlapping innovations speak distinct languages, but given the generally greater ease with which innovations diffuse between dialects as opposed to languages, it appears likely that the loss of final vowels and loss of  $*k$  are changes that diffused among these communities when they still formed points on a dialect chain.

Despite the presence of overlapping innovations in both groups, it is noteworthy that SHWNG has in general been given a better reception than CMP. Perhaps this is because the area covered by SHWNG languages is considerably smaller than that covered by CMP languages, and most lexical innovations either appeared while the languages were a single geographically compact community, or later spread throughout the shorter dialect chain that formed as a result of fairly rapid migration. By contrast, the CMP languages are distributed from the central Moluccas to the western Lesser Sundas, apparently as the result of a rapid population expansion from the region of the northern Moluccas during the initial AN penetration of eastern Indonesia.

It is equally noteworthy that the higher-level subgroups of western Oceanic proposed in Ross (1988) have received little of the critical reaction that has so far been directed at the CMP and CEMP hypotheses. This is puzzling, since Ross himself is quite candid about the overlapping nature of the innovations taken to support some of his major groupings. To some extent, this is concealed by the terminology used. Ross (1988:25) assigned all of the Oceanic languages of western Melanesia to one of four clusters, where *cluster* is defined as “a large group of languages which appear to have some kind of genetic links with each other at a level lower than that of the Oceanic subgroup itself.” These four groups are: (1) the North New Guinea cluster, (2) the Papuan Tip cluster, (3) the Meso-Melanesian cluster, and (4) the Admiralties cluster. While the Papuan Tip and Admiralties clusters are relatively noncontroversial, innovations defining the North New Guinea and Meso-Melanesian clusters share many of the same distributional traits as those defining CMP. Despite his terminological usage, Ross (1988:120) notes of the three main components of the North New Guinea cluster (the Schouten chain, the Huon Gulf family, and the Ngero/Vitiaz family) that “there are features which join these groups into a rather complicated linkage. This linkage is probably not descended from a single proto language in the conventional sense of that term, but from a network of dialects which became more or less separated from other communalects of an Oceanic linkage. . . . In other words, Proto North New Guinea was not a single communalect which diffused or split, but a linkage of communalects which, so to speak, became Proto North New Guinea at the time it became independent of other Oceanic communalects.” Similarly, although he points out (1988:261) that “there are certain innovations common to all Meso-Melanesian groups,” Ross describes the Meso-Melanesian cluster as “a portion of an early Oceanic linkage of communalects (and perhaps the eastern part of the same linkage as the early North New Guinea linkage broke off from),” thus suggesting that many of these innovations may be products of diffusion along a dialect chain. Finally, Ross (1988:346ff.) proposed that all of the Oceanic languages of western Melanesia form a larger group that he called “Western Oceanic.” However, he quickly qualified this with the observation (1988:351) that “there are no phonological innovations shared by all

Western Oceanic languages except the merger of POC \*d and \*dr,” an innovation that is also found in the Nuclear Micronesian and Southeast Solomonic languages, and in various languages of northern and central Vanuatu. I cite these cases to remind the reader that, although the CMP hypothesis must stand or fall on its own merits, it is only one of a number of subgrouping hypotheses about AN languages that depend on a complex array of overlapping innovations.

Blust (1993) proposed four phonological innovations for CMP, which D&G (2008:120) somewhat arbitrarily split into six. These are: (1) loss of prepenultimate initial syllables that began with a vowel or a laryngeal consonant (PMP \*q or \*h), (2) diphthong (glide) truncation, (3) postnasal voicing of stops, and (4) the irregular change PMP \*baqeRu to PCMP \*bəʔəRu ‘new’. Each of these will be considered in turn.

#### 4.1 LOSS OF PREPENULTIMATE INITIAL SYLLABLES THAT BEGAN WITH A VOWEL OR LARYNGEAL CONSONANT.

The history of stress in AN languages is still imperfectly understood. Proto-Philippines had a phonemic accent that has not been convincingly explained as either the continuation of an earlier pattern of phonemic stress or of conditioned change. In languages with contrastive stress, all unstressed vowels except schwa (PAN \*e) retain their full length and coloration. However, where stress is predictably penultimate, as in the southern Philippines and most AN languages elsewhere, there is a tendency for prepenultimate vowels to weaken to schwa, and in many cases to drop if they originally were, or came to be, word-initial. In every known language of Borneo for which adequate descriptive data are available, for example, PMP \*qasawa has reduced to a reflex of \*sawa ‘spouse’ (from earlier \*əsawa), and PMP \*qateluR has reduced to reflexes of \*teluR ‘egg’ (from earlier \*ətəluR). A similar tendency is widespread in the languages of eastern Indonesia, but is virtually absent in Oceanic languages.

As noted in Blust (1993:264), although nearly all reflexes of PMP \*qateluR ‘egg’ in eastern Indonesia show loss of the first syllable, Watubela *kathu* ‘egg’ preserves it. D&G (2008:120) seize on this observation as evidence that aphesis has been recurrent in the languages of eastern Indonesia, and therefore cannot carry weight as subgrouping evidence. This criticism is valid, but as just stated, it was initially raised in Blust (1993). Watubela is exceptional in preserving a nonzero reflex of PMP \*q-, which regularly became *k* in all positions, as in \*qulu > *kulu* ‘head’, \*ma-tuqah > *mtuka* ‘old’, or \*daRaQ > *lalak* ‘blood’. For this reason this language retained initial syllables of trisyllabic words that reflect PMP \*qV-. However, as noted in Blust (1993), this does not eliminate the possibility that PMP trisyllables that began with a vowel lost the initial syllable as the result of a single change, since no language in eastern Indonesia is known to preserve a prepenultimate initial vowel (cf. reflexes of \*abuqaj ‘beetle’, \*adamay ‘a plant: *Pipturus argenteus*’, \*aluten ‘firebrand’, \*anaduq ‘long, of objects’, \*aniṅal ‘echo’, \*anipa ‘large snake sp.’, \*anuliṅ ‘a fruit tree: *Pisonia umbellifera*’, and \*apuni ‘tree fern’ in Blust 1995).<sup>14</sup>

14. D&G (2008:120) cite PCMP \*anaduq ‘long’ as a PCMP reconstruction with a prepenultimate initial vowel, but this was a typographical error for \*naduq in the original source. To these eight examples we might also add reflexes of PMP \*habaRat ‘west monsoon’, \*hadiRi ‘housepost’, \*hapejes/hapejis ‘sting, smart’, and \*haRezan ‘notched log ladder’, since it is unclear whether Soboyo, which often preserves the PMP contrast of \*V- vs. \*hV- in disyllables, also preserves it in trisyllables.

**4.2 GLIDE TRUNCATION.** The PAN diphthongs \*-ay, \*-aw, \*-uy, and \*-iw were monophthongized in many daughter languages, but normally as -e, -o, -i, and -i, respectively—that is, by the equivalent of “guna assimilation.”<sup>15</sup> In two geographically separated regions—the western plains of Taiwan (including Thao, on the western slopes of the Central Mountains) and eastern Indonesia—diphthongs were instead monophthongized by truncation of the coda. Blust (1993:265) was able to access usable data only for \*-ay, \*-aw, and \*-uy, but the available examples showed that this process is irregular in many languages. While languages such as Bimanese and Komodo show “normal” monophthongization for all three diphthongs, Manggarai, for example, shows \*-aw > o and \*-uy > i, but usually shows \*-ay > a. Neighboring Ngadha, on the other hand, usually reflects both \*-ay and \*-aw as -a, but \*-uy as either -u or -i. As noted in Blust (1993), many instances of glide truncation in CMP languages must be due to diffusion. D&G (2008) interpret this unconditionally as support for an argument against CMP. However, the whole point of the original presentation, which recognized CMP as a linkage rather than a discrete subgroup such as CEMP, is that dialect chains create the conditions necessary for exuberant diffusion of innovations. While diffusion between distinct languages is possible, diffusion of this magnitude is more commonly seen in dialect networks or chains, and the observed distribution of this very distinctive change thus suggests that at least many of the languages that have been assigned to CMP were once part of an extensive dialect chain extending through the Lesser Sundas and into the southern and central Moluccas.

D&G (2008:114) claim that “the Central Malayo-Polynesian (CMP)-area languages do not convincingly meet the criteria commonly accepted for a subgroup or even a linkage.” They then elaborate on this as follows (2008:115): “The fact that Blust claims to have reconstructed a linkage, rather than an undifferentiated protolanguage, means that we should expect some degree of incompleteness in the attestation of these forms, but that the incompleteness should be geographically defined, rather than random.” Although they are nonspecific, the target of this criticism evidently is the distribution of forms that show diphthong truncation, since this change is geographically discontinuous, affecting a number of languages in the central Moluccas, then completely skipping Kédang, Sika, and the languages of Timor and Roti, before appearing again among the languages of central and western Flores. A distribution of this kind is, admittedly, puzzling. But if diphthong truncation were attributed to independent invention without contact, it would be even more puzzling that this change would be so frequent in the languages of eastern Indonesia, when it is nearly absent elsewhere in the AN language family. Some type of stimulus diffusion would appear to offer the most reasonable explanation for the observed distribution, even if its discontinuous pattern remains puzzling.<sup>16</sup>

15. Sanskritists have long used the term “guna assimilation” for a liaison-like process in Sanskrit whereby *a + i* fused to *e*, and *a + u* fused to *o*.

16. D&G (2008:124) argue that there is no reason to consider \*-ay > a, \*-aw > a, and \*-uy > i/u as part of a single sound change, since some languages show an asymmetric development. However, apart from some exceptions in individual lexical items, the only languages that truncate one of these diphthongs but not the other are Manggarai of western Flores (\*-ay > a, but \*-aw > o), and members of the Leti-Luangic group east of Timor (\*-ay > Leti *i*, \*-aw > Leti *a*). Virtually all languages of the central Moluccas, and Florinese languages such as Ngadha and Lio, truncate both diphthongs indiscriminately. If these changes were in fact independent, an explanation would need to be found for why they are associated in so many languages.

Finally, D&G (2008:124) claim that glide truncation is “widely attested in WMP.” This is simply untrue. While a number of the languages of Sulawesi reflect \*-uy as -u, for example, there is no clear evidence that this is a product of truncation rather than fusion of nucleus and coda. The changes \*-ay > -a and \*-aw > -a are far more distinctive, and neither of these has been reported in any AN language outside of eastern Indonesia and the Western Plains group in Taiwan (Blust 2003:32).

**4.3 POSTNASAL VOICING OF STOPS.** As noted in Blust (2008:94–98), one of the most striking developments in a number of the languages of the Lesser Sundas and central Moluccas is a three-step change that involves (1) loss of the vowel in PMP \*ma- ‘stative’, (2) postnasal voicing of a base-initial voiceless stop, and (3) loss of the conditioning nasal. Examples are seen in PMP \*ma-panas > Ngadha *bana*, Kodi *banaho*, Galoli *banas* ‘hot’, Buru *bana* ‘fire’; \*ma-penuq > Ngadha, Sika, Kemak *benu*, Mambai *beun* ‘full, of a container’, Buru *benu-k* (vt) ‘tamp in’; and \*ma-putiq > Kemak *buci*, Mambai, Galoli *buti*, Buru *boti* ‘white’. A fuller consideration of the comparative evidence shows that some languages preserve a prenasalized stop in various of these forms, as with Kambera *mbanahu* ‘hot’, or Ende *mbenu*, Kodi *mbanuka*, Kambera *mbinu* ‘full’. Still other languages reflect \*ma-p as *m-*, as with \*ma-panas > Tetun *manas* ‘hot’, or \*ma-putiq > Atoni *muti*, Tetun *muti-n* ‘white’. It is clear, then, that only the first two steps are common to all languages of the region. D&G (2008:125) regard these shared changes as natural products of postnasal devoicing, common to many of the world’s languages, and so dismiss them as evidence for subgrouping. What this criticism misses is that postnasal voicing involved the prior loss of the vowel in \*ma- ‘stative’, and voicing of a *stem-initial* voiceless stop, a highly distinctive development not known to be shared with any language outside eastern Indonesia.

It is clear that this change, like glide truncation, was a product of diffusion, since some languages that show no evidence of it are interspersed with those that do. Nonetheless, it is an innovation, and one that is highly distinctive. If the attested distribution is due to diffusion, this in itself implies the earlier existence of a “diffusion corridor” spanning much of eastern Indonesia, a condition that—as already noted—would be far more likely to obtain within a dialect chain than within a chain of mutually unintelligible languages. To simply dismiss it because postnasal voicing is a natural phonological process loses sight of its complex origin (syncope of a prefixal vowel, followed by postnasal voicing of a stem-initial stop in cognate forms).

**4.4 PMP \*baqeru > PCMP \*bəʔəru ‘NEW’.** Blust (1993:266) noted that a number of the languages of eastern Indonesia reflect PMP \*baqeRu ‘new’ with a penultimate schwa or a vowel that has developed from penultimate schwa. D&G (2008:127–30) correctly observe that this development could simply reflect the merger of PMP \*a and \*e in the antepenultimate syllable, followed by loss of \*q and contraction of like vowels. I concede the point, and abandon this comparison for purposes of defending the CMP hypothesis.

**4.5 SUMMARY OF THE EVIDENCE FOR CMP.** Table 4 summarizes the preceding discussion, comparing table 1 in D&G (2008:120) with the conclusions

reached here. D&G (2008:120) are concerned primarily with assessing whether the proposed innovations defining CMP are inclusive. But this has never been an issue, since CMP was explicitly described as a linkage in Blust (1993). Where the two interpretations diverge is as follows:

(1) D&G do not regard loss of prepenultimate initial vowels as having any subgrouping value for CMP. While this is a change that admittedly occurs elsewhere (as throughout Borneo), and while it clearly did not occur in Watubela of the central Moluccas, all other languages of eastern Indonesia show loss of prepenultimate syllables that began with a vowel, \*h, or \*q in PMP. This agreement may be a product of massive drift, but we cannot completely exclude the possibility that prepenultimate initial vowels had already been lost before the disappearance of \*q-

(2) D&G dismiss glide truncation as without subgrouping value, while the view adopted here is that the clear evidence for borrowing of sound changes that glide truncation presents is more nearly compatible with diffusion through a prehistoric dialect chain than through a chain of distinct languages.

(3) D&G dismiss postnasal voicing as a product of language universals, and hence convergence, while the view adopted here is that this oversimplifies the facts, which required previous syncope of a prefixal vowel followed by postnasal voicing of a stem-initial consonant. While it is not impossible that this sequence of changes could be a product of convergence, it is unlikely unless these changes were diffusing through a network of dialects.

(4) It is agreed that this innovation (PMP \*baqeRu > PCEMP \*baqəRu) is of little value for subgrouping.

**5. CONCLUSIONS.** The position of the languages of eastern Indonesia is a problem that long resisted a solution rooted in the Comparative Method of linguistics. The Central-Eastern Malayo-Polynesian hypothesis in particular represents a triumph of historical inference based on exclusively shared innovations. Moreover, it is one that draws on resources unique to one part of the world, namely a major biogeographical boundary that was crossed by a late Neolithic population that continued to expand eastward to the farthest inhabitable reaches of the Pacific. The distribution of cognate sets reflecting such terms as \*kandoRa ‘cuscus’ and \*mans(aə)r ‘bandicoot’ is a precious relic of patterns of prehistoric population split and cannot or should not be casually swept under the carpet in academic debate.

**TABLE 4. THE CMP INNOVATIONS RECONSIDERED AGAIN**

	<b>D&amp;G (2008)</b>	<b>Blust (2009)</b>
<b>Regular phonological innovations</b>		
1. Loss of initial V3	no	weak
2. Glide truncation	no	useful
3. Postnasal voicing	no	useful
<b>Irregular phonological innovations</b>		
4. PMP *baqeRu > PCEMP *baqəRu	no	not useful

“The position of the languages of eastern Indonesia and East Timor” is an ambitious paper—potentially an important paper—but it is one that is flawed both by haste and by hubris.<sup>17</sup> Neither author is a historical linguist, and they share fundamental misconceptions about the nature of innovation, upon which all reliable subgrouping hypotheses must rest. As a consequence, much valid subgrouping evidence is dismissed with little more than a wave of the hand. Some of the errors that D&G have committed perhaps could have been avoided had they consulted the Austronesian Comparative Dictionary and understood the conventions adopted there. But even if they had, their lack of training in historical linguistics probably would have led them into serious error in dealing with some of the most complex and challenging subgrouping issues that are found in this enormous and structurally diverse language family.

## REFERENCES

- Adelaar, K. A. 1992. *Proto-Malayo: The reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.
- . 2005. The Austronesian languages of Asia and Madagascar: A historical perspective. In *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 1–42. London: Routledge.
- Anceaux, J. C. 1961. *The linguistic situation in the islands of Yapen, Kurudu, Nau and Miosnum, New Guinea*. VKI 35. The Hague: Nijhoff.
- Arndt, Paul. 1961. *Wörterbuch der Ngadhasprache*. Studia Instituti Anthropos, vol. 15. Posieux, Fribourg, Switzerland: Anthropos Institute.
- Bender, Byron W., Ward H. Goodenough, Frederick H. Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Ho-Min Sohn, Stephen Trussel, and Judith W. Wang. 2003. Proto-Micronesian reconstructions—2. *Oceanic Linguistics* 42:271–358.
- Blust, Robert. 1970. Proto-Austronesian addenda. *Oceanic Linguistics* 9:104–62.
- . 1974. Eastern Austronesian: A note. *University of Hawaii Working Papers in Linguistics* 6.4:101–7. Honolulu: Department of Linguistics, University of Hawai‘i.
- . 1978a. Eastern Malayo-Polynesian: A subgrouping argument. In *Second International Conference on Austronesian Linguistics: Proceedings*, fascicle 1, ed. by S. A. Wurm and Lois Carrington, 181–234. Canberra: Pacific Linguistics.
- . 1978b. *The Proto-Oceanic palatals*. Memoir 43. Wellington: The Polynesian Society (Incorporated).
- . 1981. The Soboyo reflexes of Proto-Austronesian \*S. In *Historical linguistics in Indonesia, Part I*, ed. by Robert A. Blust, 21–30. NUSA 10. Jakarta: Atma Jaya University.
- . 1982. The linguistic value of the Wallace Line. *Bijdragen tot de taal-, land- en volkenkunde* 138:231–50.
- . 1983–84. More on the position of the languages of eastern Indonesia. *Oceanic Linguistics* 22/23:1–28.

17. The original version that I saw was riddled with factual errors, many of them quite surprising. In large part due to my feedback, most of these were eliminated. One that survived is the citation of Yami and Iralalay as distinct languages in D&G’s Map 1. In addition, D&G show little control of the comparative evidence, citing nonexistent reconstructions such as PAN \*Suai ‘younger sibling’ (121), imagining that \*qalejaw > Tagalog *araw* shows loss of an “antepenultimate laryngeal syllable” rather than medial schwa syncope followed by cluster reduction (122), or proposing etymologies such as \*kepkep ‘grasp’ > *gepe* ‘clamp’, or \*likud > *kuri* ‘behind’, despite the need to appeal to otherwise unknown phonological developments.

- . 1988. *Austronesian root theory: An essay on the limits of morphology*. Studies in Language Companion Series 19. Amsterdam/Philadelphia: John Benjamins.
- . 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241–93.
- . 1995. *Austronesian comparative dictionary*. Unfinished online comparative dictionary, available upon request.
- . 1996. The Neogrammarian hypothesis and pandemic irregularity. In *The Comparative Method reviewed: Regularity and irregularity in language change*, ed. by Mark Durie and Malcolm Ross, 135–56. New York and Oxford: Oxford University Press.
- . 2002. The history of faunal terms in Austronesian languages. *Oceanic Linguistics* 41:89–139.
- . 2003. *Thao dictionary*. Language and Linguistics Monograph Series No. A5. Academia Sinica: Institute of Linguistics (Preparatory Office).
- . 2008. Is there a Bima-Sumba subgroup? *Oceanic Linguistics* 47:45–113.
- . Forthcoming. *The Austronesian languages*. Canberra: Pacific Linguistics.
- Böhm, Karl. 1975. *Das Leben einiger Inselvölker Neuguineas*. Collectanea Instituti Anthropos 6. St. Augustin: Anthropos Institute.
- Campbell, Lyle. 2004. *Historical linguistics: An introduction*. 2nd ed. Cambridge, MA: The MIT Press.
- Capell, A. 1968 [1941]. *A new Fijian dictionary*. 3rd ed. Suva: Government Printer.
- Chowning, Ann. 1991. Proto Oceanic culture: The evidence from Melanesia. In *Currents in Pacific linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, ed. by Robert Blust, 43–75. Canberra: Pacific Linguistics.
- Collins, James T. 1982. Linguistic research in Maluku: A report of recent field work. *Oceanic Linguistics* 21:73–146.
- Darlington, Philip J., Jr. 1980 [1957]. *Zoogeography: The geographical distribution of animals*. Huntington, New York: Robert E. Krieger.
- Dempwolff, Otto. 1920. *Die Lautensprechungen der indonesischen Lippenlaute in einigen anderen austronesischen Südseesprachen*. Zeitschrift für Eingeborenen-Sprachen, Supplement 2. Berlin: Reimer.
- . 1934–38. *Vergleichende Lautlehre des austronesischen Wortschatzes*. ZfES, Supplement 1. *Induktiver Aufbau einer indonesischen Ursprache* (1934), Supplement 2. *Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen* (1937), Supplement 3. *Austronesisches Wörterverzeichnis* (1938). Berlin: Reimer.
- Devin, Chaumont. 1989. A dictionary and grammatical notes on the language of Buru island. Ms., Honolulu, June 6, 1989.
- Donohue, Mark. 2004. The pretenders to the Muna-Buton group. In *Papers in Austronesian subgrouping and dialectology*, ed. by John Bowden and Nikolaus P. Himmelmann, 21–35. Canberra: Pacific Linguistics.
- Donohue, Mark, and Charles E. Grimes. 2008. Yet more on the position of the languages of eastern Indonesia and East Timor. *Oceanic Linguistics* 47:114–58.
- Dyen, Isidore. 1965. *A lexicostatistical classification of the Austronesian languages*. Indiana University Publications in Anthropology and Linguistics, and Memoir 19 of the International Journal of American Linguistics. Baltimore: The Waverly Press.
- . 1978. The position of the languages of eastern Indonesia. In *Second International Conference on Austronesian Linguistics: Proceedings*, fascicle 1, ed. by S. A. Wurm and Lois Carrington, 235–54. Canberra: Pacific Linguistics.
- Fox, C. E. 1955. *A dictionary of the Nggela language (Florida, British Solomon Islands)*. Auckland: The Unity Press.
- Gordon, Raymond G., Jr. 2005. *Ethnologue: Languages of the world*. 15th ed. Dallas, Texas: SIL International.
- Greenhill, Simon J., Robert Blust, and Russell D. Gray. 2005–08. The Austronesian Basic Vocabulary Database. <http://language.psy.auckland.ac.nz/austronesian>.

- Grimes, Charles E. 1991. Central Malayo-Polynesian: An interim evaluation. Paper presented at the Sixth International Conference on Austronesian Linguistics, Honolulu, Hawai'i, May 20–24.
- Grimes, Charles E., Tom Therik, Barbara Dix Grimes, and Max Jacob. 1997. *A guide to the people and languages of Nusa Tenggara. Paradigma B-1*. Kupang: Artha Wacana Press.
- Jonker, J. C. G. 1908. *Rottineesch–Hollandsch woordenboek*. Leiden: E. J. Brill.
- Leplin, Jarrett. 1975. The concept of an *ad hoc* hypothesis. *Studies in history and philosophy of science* 5:309–45.
- Lynch, John. 2002. The Proto-Oceanic labiovelars: Some new observations. *Oceanic Linguistics* 41:310–62.
- Maan, G. 1940. *Boelisch–Nederlandsche woordenlijst*. Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen, 74.3. Proceedings of the Batavia Society of Arts and Sciences. Bandung: A. C. Nix & Co.
- Madulid, Domingo A. 2001. *A dictionary of Philippine plant names*. 2 vols. Makati City: Bookmark.
- Mahdi, Waruno. 1994. Some Austronesian maverick protoforms with culture-historical implications—II. *Oceanic Linguistics* 33:431–90.
- Marck, Jeff. 2000. *Topics in Polynesian language and culture history*. Canberra: Pacific Linguistics.
- Mathijssen, A. 1906. *Tettum–Hollandsche woordenlijst*. Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen, vol. 56. Batavia and The Hague: Albrecht/Nijhoff.
- Mead, David. 1998. *Proto-Bungku-Tolaki: Reconstruction of its phonology and aspects of its morphosyntax*. Ann Arbor, MI: University Microfilms International.
- Meyer, Frans. 1964. *Sikkaneesch–Nederlands woordenboek*. Typescript.
- Milke, Wilhelm. 1968. Proto-Oceanic addenda. *Oceanic Linguistics* 7:147–71.
- Mills, Roger F. 1975. *Proto South Sulawesi and Proto Austronesian phonology*. 2 vols. PhD diss., Department of Linguistics, The University of Michigan. Ann Arbor, MI: University Microfilms International.
- Moeliono, Anton M., et al. 1989. 2nd printing. *Kamus besar Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Nothofer, Bernd. 1992. CEMP and CMP—two Austronesian subgroups? A critical evaluation of Blust (1990) and Grimes (1991). Paper presented at the Second International Maluku Studies Conference, Honolulu, Hawai'i, July 30–August 2.
- Pawley, Andrew, and Timoci Sayaba. 2003. *Words of Waya: A dictionary of the Wayan dialect of the Western Fijian language*. 2 vols. Draft.
- Popper, Karl R. 1968 [1935]. *The logic of scientific discovery*. Harper Torchbooks. New York: Harper and Row.
- Ross, Malcolm D. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.
- . 1995. Some current issues in Austronesian linguistics. In *Comparative Austronesian dictionary: An introduction to Austronesian studies. Part 1*, fascicle 1, ed. by Darrell T. Tryon, 45–120. Berlin: Mouton de Gruyter.
- . 1997. Social networks and kinds of speech-community event. In *Archaeology and Language I: Theoretical and methodological orientations*, ed. by Roger Blench and Matthew Spriggs, 209–61. London: Routledge.
- Ross, Malcolm, Andrew Pawley, and Meredith Osmond, eds. 1998. *The lexicon of Proto Oceanic, the culture and environment of ancestral Oceanic society*, vol. 1: *Material culture*. Canberra: Pacific Linguistics.
- . 2003. *The lexicon of Proto Oceanic, the culture and environment of ancestral Oceanic society*, vol. 2: *The physical environment*. Canberra: Pacific Linguistics.

- Ruhlen, Merritt. 1987. *A guide to the world's languages*, vol. 1: *Classification*. Stanford, CA: Stanford University Press.
- Samely, Ursula. 1991. *Kédang (eastern Indonesia), some aspects of its grammar*. Hamburg: H. Buske.
- Sneddon, J. N. 1978. *Proto-Minahasan: Phonology, morphology and wordlist*. Canberra: Pacific Linguistics.
- . 1984. *Proto-Sangiric and the Sangiric languages*. Canberra: Pacific Linguistics.
- Stokhof, W. A. L., ed. 1982. *Holle lists: Vocabularies in languages of Indonesia*, vol. 3/3. Canberra: Pacific Linguistics.
- Stresemann, E. 1927. *Die Lauterscheinungen in den ambonischen Sprachen*. Zeitschrift für Eingeborenen-Sprachen, Supplement 10. Berlin: Reimer.
- Tryon, Darrell T., ed. 1995. *Comparative Austronesian dictionary: An introduction to Austronesian studies*. 4 vols. Berlin: Mouton de Gruyter.
- Tryon, D. T., and B. D. Hackman. 1983. *Solomon Islands languages: An internal classification*. Canberra: Pacific Linguistics.
- Tsuchida, Shigeru, Yukihiro Yamada, and Tsunekazu Moriguchi. 1987. *Lists of selected words of Batanic languages*. Tokyo: Department of Linguistics, Faculty of Letters, University of Tokyo.
- Usup, Hunggu Tajudin. 1981. *Rekonstruksi fonem proto kelompok bahasa Gorontalo sebelah timur*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
- van den Berg, René. 1991. Muna historical phonology. In *Studies in Sulawesi Linguistics, Part II*, ed. by J. N. Sneddon, 2–28. NUSA, vol. 33. Jakarta: Universitas Katolik Indonesia Atma Jaya.
- Wallace, Alfred Russel. 1962 [1869]. *The Malay archipelago*. New York: Dover Publications, Inc.
- Walsh, D. S., and Bruce Biggs. 1966. *Proto-Polynesian word list I*. Te Reo Monographs. Auckland: Linguistic Society of New Zealand.
- Wilkinson, R. J. 1959. 2 vols. *A Malay-English dictionary (romanised)*. London: Macmillan & Co., Ltd.
- Wolff, John U. 1996. The development of the passive verb with pronominal prefix in Western Austronesian languages. In *Reconstruction, classification, description—Festschrift in honor of Isidore Dyen*, ed. by Bernd Nothofer, 15–40. Hamburg: Abera.